

21 世纪大学本科计算机专业系列教材

计算机组成原理教师用书

蒋本珊 编著

清华大学出版社
北京

内 容 简 介

本书是与“21 世纪大学本科计算机专业系列教材”中的《计算机组成原理》(主教材)一书完全配套的教师用书。全书共分 8 章,与主教材的结构相同,每一章都按基本内容要求、误点疑点解惑、相关知识介绍和教材习题解答四大版块进行组织。

全书概念清楚、通俗易懂、由浅入深,其核心内容是每一章的误点疑点解惑和相关知识介绍两大版块。各章中都以专题的形式对有关问题进行了比较详细和深入的讨论,并且通过一些例题来帮助读者加深对“计算机组成原理”课程所学知识的理解。教材习题解答版块则给出了主教材中所附全部习题的详细解答过程和参考答案。

本书是教师讲授“计算机组成原理”课程的教学参考书,也可以作为学生学习本课程的参考用书。

版权所有,翻印必究。举报电话:010-62782989 13501256678 13801310933

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

本书防伪标签采用特殊防伪技术,用户可通过在图案表面涂抹清水,图案消失,水干后图案复现;或将表面膜揭下,放在白纸上用彩笔涂抹,图案在白纸上再现的方法识别真伪。

图书在版编目(CIP)数据

计算机组成原理教师用书/蒋本珊编著. —北京:清华大学出版社,2005.8

(21 世纪大学本科计算机专业系列教材)

ISBN 7-302-11316-5

I. 计… II. 蒋… III. 计算机体系结构—高等学校—教学参考资料 IV. TP303

中国版本图书馆 CIP 数据核字(2005)第 074701 号

出 版 者:清华大学出版社

<http://www.tup.com.cn>

社 总 机:010-62770175

地 址:北京清华大学学研大厦

邮 编:100084

客户服务:010-62776969

责任编辑:张瑞庆

封面设计:孟繁聪

印 装 者:北京鑫海金澳胶印有限公司

发 行 者:新华书店总店北京发行所

开 本:185×230 印张:16.5 字数:346 千字

版 次:2005 年 8 月第 1 版 2006 年 7 月第 2 次印刷

书 号:ISBN 7-302-11316-5/TP·7456

印 数:3001~

定 价:25.00 元



21 世纪大学本科计算机专业系列教材编委会

名誉主任：陈火旺

主 任：李晓明

副 主 任：钱德沛 焦金生

委 员：(按姓氏笔画为序)

马殿富	王志英	王晓东	宁 洪	刘 辰
孙茂松	李大友	李仲麟	吴朝晖	何炎祥
宋方敏	张大方	张长海	周兴社	侯文永
袁开榜	钱乐秋	黄国兴	蒋宗礼	曾 明
廖明宏	樊孝忠			

秘 书：张瑞庆

本书责任编辑：袁开榜

序 言

PREFACE

21 世纪是知识经济的时代,是人才竞争的时代。随着 21 世纪的到来,人类已步入信息社会,信息产业正成为全球经济的主导产业。计算机科学与技术的信息产业中占据了最重要的地位,这就对培养 21 世纪高素质创新型计算机专业人才提出了迫切的要求。

为了培养高素质创新型人才,必须建立高水平的教学计划和课程体系。在 20 多年跟踪分析 ACM 和 IEEE 计算机课程体系的基础上,紧跟计算机科学与技术的发展潮流,及时制定并修正教学计划和课程体系是尤其重要的。计算机科学与技术的发展对高水平人才的要求,需要我们从总体上优化课程结构,精炼教学内容,拓宽专业基础,加强教学实践,特别注重综合素质的培养,形成“基础课程精深,专业课程宽新”的格局。

为了适应计算机科学与技术学科发展和计算机教学计划的需要,要采取多种措施鼓励长期从事计算机教学和科技前沿研究的专家教授积极参与计算机专业教材的编著和更新,在教材中及时反映学科前沿的研究成果与发展趋势,以高水平的科研促进教材建设。同时适当引进国外先进的原版教材。

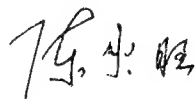
为了提高教学质量,需要不断改革教学方法与手段,倡导因材施教,强调知识的总结、梳理、推演和挖掘,通过加快教案的不断更新,使学生掌握教材中未及时反映的学科发展新动向,进一步拓宽视野。教学与科研相结合是培养学生实践能力的有效途径。高水平的科研可以为教学提供最先进的高新技术平台和创造性的工作环境,使学生得以接触最先进的计算机理论、技术和环境。高水平的科研还可以为高水平人才的素质教育提供良好的物质基础。学生在课题研究中不但能了解科学研究的艰辛和科研工作者的奉献精神,而且能熏陶和培养良好的科研作风,锻炼和培养攻关能力和协作精神。

进入 21 世纪,我国高等教育进入了前所未有的大发展时期,时代的进步与发展对高等教育质量提出了更高、更新的要求。2001 年 8 月,教育部颁发了《关于加强高等学校本科教学工作,提高教学质量的若干意见》。文件指出,本科教育是高等教育的主体和基础,抓好本科教学是提高整个高等教育质量的重点和关键。随着高等教育的普及和高等学校的扩招,在校大学本科计算机专业学生的人数将大量上升,对适合 21 世纪大学本科计算机科学与技术学科课程体系要求的,并且适合中国学生学习的计算机专业教材的需求量

也将急剧增加。为此,中国计算机学会和清华大学出版社共同规划了面向全国高等院校计算机专业本科生的“**21 世纪大学本科计算机专业系列教材**”。本系列教材借鉴美国 ACM 和 IEEE/CS 最新制定的 *Computing Curricula 2001*(简称 CC2001)课程体系,反映当代计算机科学与技术学科水平和计算机科学技术的新发展、新技术,并且结合中国计算机教育改革成果和中国国情。

中国计算机学会教育专业委员会和全国高等学校计算机教育研究会,在清华大学出版社的大力支持下,跟踪分析 CC2001,并结合中国计算机科学与技术学科的发展现状和计算机教育的改革成果,研究出了《中国计算机科学与技术学科教程 2002》(*China Computing Curricula 2002*,简称 CCC2002),该项研究成果对中国高等学校计算机科学与技术学科教育的改革和发展具有重要的参考价值和积极的推动作用。

“**21 世纪大学本科计算机专业系列教材**”正是借鉴美国 ACM 和 IEEE/CS CC2001 课程体系,依据 CCC2002 基本要求组织编写的计算机专业教材。相信通过这套教材的编写和出版,能够在内容和形式上显著地提高我国计算机专业教材的整体水平,继而提高我国大学本科计算机专业的教学质量,培养出符合时代发展要求的具有较强国际竞争力的高素质创新型计算机人才。



中国工程院院士

国防科学技术大学教授

21 世纪大学本科计算机专业系列教材编委会名誉主任

2002 年 7 月

前 言

FOREWORD

“计算机组成原理”是计算机类专业学生的必修核心课程之一,主要讨论计算机各部件的基本组成原理,以及各部件互连构成整机系统的技术。本课程在计算机科学与技术学科中处于承上启下的地位,具有内容多、难度大等特点。本书根据作者本人二十年来从事“计算机组成原理”课程教学的经验 and 体会整理编写而成,以满足讲授“计算机组成原理”课程教师的需要。本书的使用将有助于教师对主教材和相关背景知识的理解,对于改进教学方法,提高教学质量都有着积极的意义。

本书是与“21 世纪大学本科计算机专业系列教材”中的《计算机组成原理》(主教材)一书完全配套的教师参考用书。全书共分 8 章,与主教材的结构相同,每一章都按基本内容要求、误点疑点解惑、相关知识介绍和教材习题解答等四大版块进行组织。

第一版块按照了解、理解、掌握 3 个不同的层次对各章节的教学内容提出了基本要求,既方便教师在教学过程中根据实际的教学时数合理地安排教学内容,又方便学生在学习过程中把握住重点。

第二版块结合作者多年的教学经验和体会,对本课程学习过程中容易出现的误点与疑点问题进行答疑解惑,指出了教学过程中需要特别注意的问题。

第三版块对主教材中由于篇幅原因没能展开的内容以及与本课程密切相关的背景知识进行介绍和讨论,以丰富读者的视野。

第四版块则给出了主教材中所附全部习题较为详细的解答过程与参考答案,这是应读者的要求而编写的。

本书是根据中国计算机学会教育委员会制订的《中国计算机科学与技术学科教程 2002》(CCC2002)对课程教学内容的要求,结合作者讲授本课程二十年的教学经验和体会“磨”出来的。全书概念清楚、由浅入深。全书的核心内容是每一章的误点疑点解惑和相关知识介绍两大版块,在每章中都以专题的形式对有关问题进行了比较详细和深入的讨论,并且还有一些例题用来帮助读者加深对有关知识点的理解。

考虑到本书的主要读者对象应该是讲授“计算机组成原理”课程的教师,所以在本书每一章的最后一个版块给出了主教材中全部习题的详细解答,以供讲授和辅导时参考。要注意的是,有些习题的答案并不惟一,设计也不一定最优,读者可以根据解题思路自己

解答,不要受到参考答案的限制和束缚。还需要特别指出的是,学生在学习过程中最好不要先看这一部分的内容,一定要给自己留下一个独立思考的空间。

“计算机组成原理”课程的教材在国内已经出版有多种,近年来也出现了一些面向学生的学习指导用书,但目前还没有见到有针对主讲和辅导教师编写的教师用书面世,本书的出版可以说是填补了一个空白,相信它会为广大讲授该课程的教师提供有益的帮助。

主教材《计算机组成原理》一书至 2004 年 3 月出版以来,受到读者的欢迎和专家的认可,并已于 2004 年底被评为北京市精品教材,与主教材配套的《计算机组成原理学习指导与习题解析》一书也已出版,此次本教师用书的出版将会使这套书更丰富和完善,它们将与《计算机组成原理电子教案》一起,构成一个“计算机组成原理”课程的立体教材教学资源体系。

本书既与主教材有紧密的关系,又独立成书,可以单独使用。既可以作为教师讲授“计算机组成原理”课程的参考书,也可以作为学生学习“计算机组成原理”课程的参考书。

在本书编写过程中得到了《21 世纪大学本科计算机专业系列教材》编委会的多次指导和建议,清华大学出版社的编辑们也为本书的出版做了许多工作。在此对他们辛勤的工作和热情的支持表示诚挚的感谢!

由于时间的原因以及个人的水平限制,书中难免有错误和不妥之处,欢迎同行专家学者和广大读者批评指正。如有问题可直接与作者邮箱联系:bsjiang@public.bta.net.cn。

作 者

2005 年 5 月于北京理工大学

目 录

CONTENTS

第 1 章 概论	1
1.1 基本要求	1
1.2 误点疑点解惑	1
1. 主机	1
2. 单总线	2
3. 完整的计算机系统	2
4. 硬件、软件的功能划分与逻辑上的等价	2
5. 机器字长和数据通路宽度	3
1.3 相关知识介绍	3
1. 冯·诺依曼型计算机及其计算机系统结构的发展	3
2. 微处理器	4
3. 三态门和总线电路	5
4. 系列机与兼容机	6
5. 计算机的多层次结构	6
6. 广义语言与计算机程序	6
7. 透明性	7
1.4 教材习题解答	8
第 2 章 数据的机器层次表示	10
2.1 基本要求	10
2.2 误点疑点解惑	11
1. 真值和机器数的区别	11
2. 模与补码表示法	11
3. 原码和补码的区别	12
4. 定点数的表示范围	12

5. 浮点数的表示范围	14
6. 浮点数的规格化	15
7. 移码偏置值的选择	15
8. 定点数与浮点数的比较	16
9. 3 种汉字编码的区别	17
10. 十进制数的 BCD 编码	18
11. 奇偶校验位的形成和奇偶校验码的检测	19
2.3 相关知识介绍	20
1. 补码 $[X]_{\text{补}}$ 与真值 X 的转换	20
2. 原码 $[X]_{\text{原}}$ 与补码 $[X]_{\text{补}}$ 的转换	21
3. 浮点数的表数范围	22
4. 浮点数尾数基数的选择	22
5. IEEE 754 标准的浮点数	24
6. 汉字的字形码	25
7. 校验码的码距	26
8. 海明编码	27
9. 循环冗余校验码的模 2 运算	28
10. 循环冗余校验码的编码和校验过程	29
11. 循环冗余校验码的纠错原理	30
2.4 教材习题解答	32
第 3 章 指令系统	41
3.1 基本内容要求	41
3.2 误点疑点解惑	42
1. 指令长度	42
2. 双操作数运算类指令的执行	42
3. 地址个数对程序长度和指令长度的影响	44
4. 不同地址数指令的进一步分析	47
5. 非规整性编码——扩展操作码法	47
6. 地址码位数与主存容量和最小寻址单位的关系	48
7. 指令寻址和数据寻址	49
8. 常见数据寻址方式分析	49
9. 各种数据寻址方式的速度比较	50
10. 变址寻址和基址寻址的区别	51

11. 相对寻址中的位移量	52
12. 存储器堆栈的操作	53
13. 返回指令的地址字段	54
14. 输入输出指令的设置	54
3.3 相关知识介绍	55
1. 操作码优化法——Huffman 编码	55
2. 操作码优化法——扩展操作码	57
3. 面向不同对象的寻址方式	59
4. 变址寻址和间接寻址的比较	59
5. 程序在主存中的定位技术	60
6. 缩短指令中地址码长度的方法	61
7. 存储器堆栈组织	61
8. 其他程序控制类指令	62
9. 对指令系统的基本要求	63
3.4 教材习题解答	64
第 4 章 数值的机器运算	72
4.1 基本内容要求	72
4.2 误点疑点解惑	73
1. 并行加法器的进位产生和传递	73
2. 并行加法器的进位传递方式和传递时间	73
3. 补码加减运算及其实现	74
4. 补码 Booth 乘法	75
5. 补码加减交替除法	75
6. 浮点加减运算中的对阶和结果规格化	76
7. 浮点除法运算中的尾数调整	77
8. BCD 码的加法运算	77
9. 基本逻辑运算及其应用	78
4.3 相关知识介绍	79
1. 全加器电路	79
2. 4 位先行进位电路 CLA	80
3. 4 位成组先行进位电路	81
4. 4 位 CLA 加法器和 4 位 BCLA 加法器的比较	81
5. 原码和反码的加减运算	81

6. 补码的移位操作	84
7. 各种舍入方法的比较	84
8. 补码校正乘法	86
9. 原码两位乘法	86
10. 不同情况除法运算中的寄存器安排	89
11. 二进制移码加减法	89
12. 浮点乘法运算的溢出和舍入问题	91
13. 多功能算术逻辑单元 74181	93
14. 先行进位发生器 74182	94
15. 位片式运算器	95
4.4 教材习题解答	96
第 5 章 存储系统和结构	108
5.1 基本内容要求	108
5.2 误点疑点解惑	109
1. 存储系统和存储器	109
2. 主存储器组织	110
3. 字节编址计算机的大端方案和小端方案	110
4. 主存储器的存储容量和存取速度	111
5. 边界对齐的数据存放方法	111
6. 动态随机存储器的刷新	112
7. 各类半导体存储芯片的特点	113
8. 存储芯片的地址译码系统	114
9. 存储容量的扩展及存储芯片与 CPU 的连接	115
10. 选片地址的全译码和部分译码	116
5.3 相关知识介绍	116
1. 存储系统的性能分析	116
2. 访问的局部性原理	117
3. 静态 RAM 芯片分析	118
4. 动态 RAM 芯片分析	119
5. MROM 和 PROM 的写入	121
6. RAM 的奇偶校验电路	122
7. BIOS 和 CMOS 芯片	122
8. 双端口存储器	123

9. 多体并行系统	123
10. 主存与 Cache 之间的地址变换与映像方式	125
11. 虚拟存储器的工作过程	125
12. 相联存储器	127
5.4 教材习题解答	128
第 6 章 中央处理器	141
6.1 基本内容要求	141
6.2 误点疑点解惑	142
1. CPU 中寄存器的设置	142
2. CPU 的主要性能指标参数	143
3. 控制器的功能与组成	144
4. 控制器的核心——控制单元	145
5. 指令的机器周期	146
6. 指令执行的控制方式	147
7. 指令微操作序列的安排	148
8. 字段直接编码和字段间接编码	153
9. 微程序控制方式	153
10. 形成后继微地址的几种方式比较	154
6.3 相关知识介绍	155
1. 外频与前端总线频率的区别与联系	155
2. CPU 的性能	156
3. 微程序控制器结构	158
4. 后继微地址形成实例	159
5. 微程序设计举例	164
6. 毫微程序设计和毫微程序控制器	166
7. 流水线的性能分析	168
8. 消除流水线瓶颈的方法	169
6.4 教材习题解答	170
第 7 章 外部设备	182
7.1 基本内容要求	182
7.2 误点疑点解惑	183
1. 磁盘存储器的平均存取时间	183

2. 改进调频制提高记录密度的分析	183
3. 硬盘存储器的圆柱面	184
4. 磁盘的基本操作	184
5. 非编码键盘的行反转法	185
6. 点阵针式打印机的打印方式	186
7. 激光打印机的印字原理	186
8. 图形和图像	187
9. CRT 显示器的有关技术指标	187
10. VRAM 的容量和内容	188
11. 字符显示原理和具体显示过程	189
12. 显示器的同步控制	190
7.3 相关知识介绍	193
1. 磁记录方式的性能特点	193
2. 几种磁记录方式的读出过程分析	195
3. 群码制(GCR)	195
4. 1/4 英寸的数据流磁带机	196
5. CD-ROM 读盘方式	196
6. 光盘刻录机工作原理	197
7. 无线鼠标	198
8. 鼓式宽行打印机	199
9. 色光三原色和颜料三原色	200
10. 彩色喷墨打印机工作原理	201
11. 彩色激光打印机色彩合成原理	202
12. AGP 接口	202
7.4 教材习题解答	203
第 8 章 输入输出系统	211
8.1 基本内容要求	211
8.2 误点疑点解惑	212
1. 接口与接口中的寄存器	212
2. 程序查询方式传送举例	213
3. 中断系统的功能	213
4. 程序中断方式不适合高速外设数据传送的原因	214
5. 中断响应阶段完成的任务	214

6. 中断服务程序入口地址的获取方式	215
7. 程序中断方式中容易混淆的几个问题	216
8. DMA 控制器的控制过程	217
9. 单字传送与成组连续传送	218
10. 周期挪用法的特点	218
11. 采用周期挪用法的数据传送过程	219
12. 3 种不同类型通道的比较	220
13. 通道操作的全过程	221
8.3 相关知识介绍	222
1. 输入输出系统的特点	222
2. 主机与外设的连接方式	223
3. 中断系统的软硬件功能分配	224
4. 可屏蔽中断和不可屏蔽中断	225
5. 向量中断与向量地址的产生	225
6. 向量中断的执行过程	227
7. 中断升级的另一种方法——改变处理机优先级	228
8. 连接多台外设的 DMA 控制器	230
9. 通道中的数据传送过程	230
10. 通道的流量分析	232
8.4 教材习题解答	234
参考文献	246

第 1 章

概 论

1.1 基本内容要求

本章将从存储程序的概念入手,讨论计算机的基本组成与工作原理,使读者对于计算机系统先有一个简单的整体概念,为今后深入讨论各个部件打下基础。

学习要求

- ◆ 了解存储程序概念
- ◆ 掌握 CPU 和主机两个术语的含义
- ◆ 掌握 5 大基本部件的功能
- ◆ 理解总线概念和总线分时共享的特点
- ◆ 理解三态门与总线电路
- ◆ 了解大、中型计算机的典型结构
- ◆ 理解计算机系统的含义
- ◆ 理解硬件与软件的关系
- ◆ 了解系列机和软件兼容
- ◆ 了解计算机系统的多层次结构
- ◆ 了解实际机器和虚拟机器
- ◆ 理解计算机中主要性能指标(基本字长、数据通路宽度、存储容量等)

1.2 误点疑点解惑

1. 主机

主机=CPU+主存储器

主机是一个简单的基本概念,但经常会有学生回答:主机=CPU+存储器。这个答案对早期的计算机来说不能算错,但对现代的计算机来说就不能算对了,起码这个答案是

不完整的。因为存储器有主存储器和辅助存储器之分,主机中只包括主存储器,而不包括辅助存储器。主存储器由 RAM 和 ROM 组成,对于微型计算机而言,是指插在主板上的内存条和其他存储芯片。辅助存储器则是硬盘、软盘、光盘等存储器的总称,它们处于主板之外,属于外部设备。

2. 单总线

总线结构是小型、微型计算机的典型结构,它可以将五大基本部件连接成硬件系统。单总线(系统总线)按总线上传送信息的不同又可以细分为:地址总线、数据总线和控制总线。地址总线用来传输由 CPU 向主存、外设发送的地址信息,其位数决定了系统能够使用的最大的存储容量;数据总线用来传输各功能部件之间的数据信息,其位数是决定系统总体性能的关键因素;控制总线上传输的是控制信息,包括 CPU 送出的控制命令和主存(或外设)返回 CPU 的反馈信号。

一提到地址总线、数据总线和控制总线,不少人可能会把它们误认为是 3 组不同的总线。事实上地址总线、数据总线和控制总线都是系统总线的一部分,只是根据总线上传送的信息不同而分别定名,不能因为它们的名称不同而认为它们是 3 个总线。

3. 完整的计算机系统

一个完整的计算机系统包含硬件系统和软件系统两大部分。

硬件系统包括运算器、控制器、存储器、输入设备和输出设备五大基本部件。

软件系统分为系统软件和应用软件两大类。系统软件包括操作系统、诊断程序、计算机语言处理程序等;应用程序包括厂家出售的通用软件 and 用户自己编写的应用程序。

这是一个简单的基本概念,但经常会有学生误认为计算机的硬件系统就是计算机系统。应当强调指出硬件和软件是相辅相成的、不可分割的整体。软件是计算机系统的灵魂,没有软件的硬件“裸机”将不能提供给用户使用,犹如一堆废铁。

4. 硬件、软件的功能划分与逻辑上的等价

硬件是躯体,是物质基础;软件是灵魂,是硬件功能的完善和补充。没有硬件,或者没有良好的硬件,就无从谈起运行软件,也就无法计算、处理某一方面的问题。没有软件,或者没有优秀的软件,计算机就是一个空壳,根本无法工作,或者不能高效率地工作。因此,硬件与软件具有相互渗透、相互依存、互相配合、互相促进的关系,二者缺一不可。

硬件与软件之间的功能分配关系常常随着技术发展而变化,哪些功能分配给硬件,哪些功能分配给软件是没有固定模式的。在计算机中,实际上有许多功能既可以直接由硬件实现,也可以在硬件的支持下依靠软件来实现,也就是说硬件和软件在逻辑功能上是等价的。例如,乘法运算,既可以用硬件乘法器实现,也可以用乘法子程序实现。在设计一台计算机时,硬、软件功能如何分配取决于所选定的设计目标、系统的性能价格比,也与当时的技术水平有关。

早期较多采用“硬件软化”的技术策略。为了降低计算机的造价,只让硬件完成比较

简单的指令操作,如传送、加法、减法、移位和基本逻辑运算,而乘法、除法、浮点运算等比较复杂的功能则交给软件完成。随着集成电路技术的飞速发展,“软件硬化”已成为常用的技术策略。将原来依靠软件才能实现的一些功能改由大规模或超大规模集成电路直接实现,如浮点运算、存储管理等。

微程序控制技术的出现使计算机结构和硬、软件功能分配发生了变化,对指令的解释与执行是通过运行微程序来实现的。因而又出现了另一种技术策略“软件固化”。利用程序设计技术可使原来属于软件级的一些功能纳入微程序一级。微程序类似于软件,但被固化在只读存储器中,属于硬件 CPU 的范畴,称为固件。人们也常采用软件固化的策略,将系统软件的核心部分(如操作系统的内核、常用软件中固定不变的部分)固化在存储芯片中。

5. 机器字长和数据通路宽度

机器字长也称基本字长,它是指参与运算的数的基本位数,也即 CPU 在同一时间内能一次处理的二进制数的位数。机器字长标志着计算精度,也反映寄存器、运算部件和数据总线的位数。机器字长越长,操作数的位数越多,计算精度也就越高,但相应部件的位数也会增多,使硬件成本随着增高。为了较好地协调计算精度与硬件成本的制约关系,针对不同需求,大多数计算机允许采用变字长运算,即允许硬件实现以字节为单位的运算以及某种基本字长或双字长的运算,通过软件实现多字长运算。

数据通路宽度是指数据总线一次所能并行传送信息的位数,它影响计算机的有效处理速度。数据通路宽度分为 CPU 内部和 CPU 外部两种情况。CPU 内部数据通路宽度一般等于机器字长,即内部数据线的位数,而 CPU 外部数据通路宽度则等于系统数据总线一次所能并行传送信息的位数,即 CPU 与主存、输入输出设备之间一次数据传送的信息位数。有的 CPU 内、外数据通路宽度一样,而有的 CPU 内、外数据通路宽度则不同。例如,Pentium 微处理器的内部数据线为 32 位,而外部数据线为 64 位。

还需要说明的一个概念是字(Word),字实际上只能算作一个计量单位,对于系列机来说,字的长度是固定的。例如,在 80x86 系列中,一个字等于 16 位。所以将 16 位的数据称为单字,32 位的数据称为双字,64 位的数据称为四倍字。在 IBM 303X 系列中,一个字等于 32 位。所以将 16 位的数据称为半字,32 位的数据称为单字,64 位的数据称为双字。

1.3 相关知识介绍

1. 冯·诺依曼型计算机及其计算机系统结构的发展

1946 年,冯·诺依曼等 3 人共同发表一篇题为“电子计算机装置逻辑结构初探”的论文,在文中详细描述了计算机的逻辑设计、指令修改的概念以及计算机的电子电路,提出了一个完整的现代计算机雏形,它由运算器、控制器、存储器和输入输出设备组成,如图

1-1 所示。

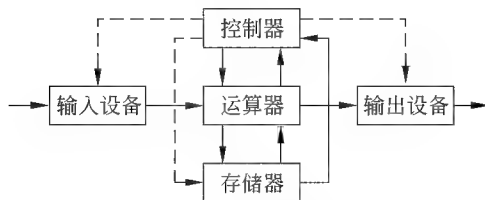


图 1-1 早期的冯·诺依曼型计算机组成框图

冯·诺依曼结构规定控制器是根据存放在存储器中的程序来工作的,即计算机的工作过程就是运行程序的过程。为了使计算机能正常工作,程序必须预先存放在存储器中。这就是存储程序的概念。

现代计算机与早期计算机相比在结构上还是有不少变化的,如从以运算器为中心改为以存储器为中心。但就其结构原理来说,目前绝大多数计算机仍建立在存储程序概念的基础上。冯·诺依曼型计算机的这种工作方式称为控制驱动。控制驱动是由指令流来驱动数据流的。

随着计算机技术的不断发展,计算机系统结构有了许多改进。主要包括:

- 从基于串行算法变为适应并行算法,出现了向量计算机、并行计算机、多处理机等。
- 高级语言与机器语言的语义距离缩小,出现了面向高级语言的计算机和直接执行高级语言的计算机。
- 硬件子系统与操作系统和数据库管理系统软件相适应,出现了面向操作系统的计算机和数据库计算机等。
- 从传统的控制驱动型改变为数据驱动型和需求驱动型,出现了数据流计算机和归约机。
- 为适应特定应用环境而出现了各种专用计算机,如快速傅里叶变换机器、过程控制计算机等。
- 为获得高可靠性而研制容错计算机。
- 计算机系统功能分散化、专业化,出现了各种功能分布计算机,包括外围处理机、通信处理机等。
- 出现了与大规模、超大规模集成电路相适应的计算机系统结构。
- 出现了处理非数值化信息的智能计算机,例如,处理自然语言、声音、图形和图像等信息的计算机。

2. 微处理器

通常将运算器和控制器合称为中央处理器(CPU),在由超大规模集成电路构成的微

型计算机中,往往将 CPU 制成一块芯片,称为微处理器。在现代的微处理器芯片中,还包含浮点处理部件(FPU)、内部高速缓冲存储器(L1 Cache)和存储管理部件(MMU),以加快计算机执行指令的速度。典型的微处理器芯片内部如图 1-2 所示。

随着集成电路技术的发展,在一些微处理器中,将 L2 Cache 也嵌入在微处理器内或将 L2 Cache 与其他部件一起封装在被称为 CPU 模块的金属盒内。

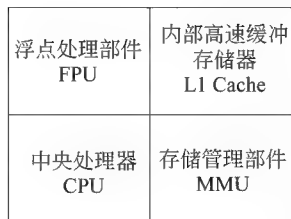


图 1-2 典型的微处理器芯片

3. 三态门和总线电路

三态门与普通门相比,多了一个控制端 G/\overline{G} ,只有当控制端有效时,该三态门才满足正常的逻辑关系,否则输出将呈现高阻状态,相当于这个三态门与外界断开联系。

三态门有多种不同的实现方式,其真值表各不相同。使用时须注意控制端是高电平有效还是低电平有效,这可以从三态门逻辑符号控制端是否标有小圆圈来区分。同样,三态门的输出也不一定反向,需从三态门逻辑符号输出端是否标有小圆圈来区分。

目前计算机中广泛采用三态门构成总线电路。总线电路有单向总线和双向总线之分,单向总线是指总线上的信息只能向一个方向传送,如地址总线就是单向总线;双向总线是指总线上的信息可以向两个方向传送,如数据总线就是双向总线。常用的总线电路有很多种,其中:总线缓冲器/驱动器/接收器是单向总线电路,总线收发器是双向总线电路。另外还有许多带有三态输出的逻辑电路,如寄存器、锁存器、数据选择器等。

74244 是一种八总线缓冲器/驱动器/接收器芯片,其内部结构如图 1-3 所示,输入输出真值表见表 1-1,其中 Z 表示高阻。当控制端 $G=0$ 时,总线输出 $Y=A$;当控制端 $\overline{G}=1$ 时,总线输出呈高阻状态。

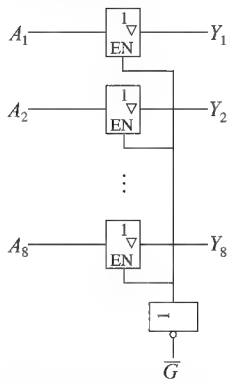


图 1-3 单向总线电路

表 1-1 74244 真值表

输入		输出
\overline{G}	A	Y
0	0	0
0	1	1
1	×	Z

74245 是一种八总线收发器,其内部结构如图 1-4 所示,输入输出真值表见表 1-2。当控制端 $G=0$ 且 $DIR=0$ 时,数据从 B 端传送到 A 端;当控制端 $G=0$ 且 $DIR=1$ 时,数据从 A 端传送到 B 端;当控制端 $G=1$ 时,总线呈高阻状态。

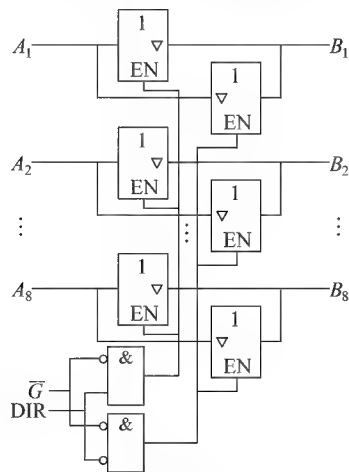


图 1 4 双向总线电路

表 1 2 74245 真值表

控制输入		操作
\overline{G}	DIR	
0	0	B 数据到 A 输出
0	1	A 数据到 B 输出
1	\times	隔离

4. 系列机与兼容机

系列机是指同一个生产厂家生产的具有相同的系统结构,但具有不同组成和实现的一系列不同型号的计算机。

兼容机是指不同生产厂家生产的具有相同系统结构的计算机。它的思想与系列机的思想是一致的。

5. 计算机的多层次结构

计算机系统由硬件、固件和软件组成,按功能划分成多级层次结构。每一级各对应一种机器,其作用和组成如图 1-5 所示。在这里,“机器”只对一定的观察者而存在。它的功能体现在广义语言上,能对该语言提供解释手段,如同一个解释器,然后作用在信息处理和控制对象上。在某一级观察者看来,他只是通过该级的语言来了解和使用计算机,至于下层机器级是如何工作和如何实现的就不必关心了。

把计算机系统按功能划分成多级层次结构,首先有利于正确理解计算机系统的工作,明确软件、硬件和固件在计算机系统中的地位 and 作用;其次有利于理解各种语言的实质及其实现;最后还有利于探索虚拟机器新的实现方法,设计新的计算机系统。

6. 广义语言与计算机程序

广义语言包括机器语言、汇编语言、高级语言和应用语言等。

机器语言(机器指令)是计算机能直接识别和执行的语言,但用机器语言编写程序、阅

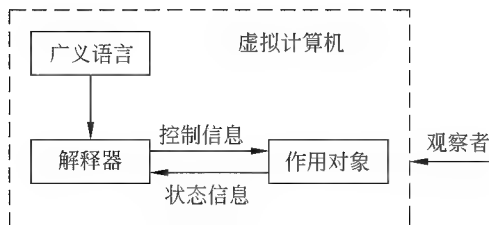


图 1-5 机器的作用和组成

读程序都非常困难。为了提高编程序、读程序的效率,产生了与机器语言相对应的符号(助记符)语言,这种符号语言后来就发展成了汇编语言。因为机器不认识汇编语言,所以必须通过叫做汇编程序的软件把它转换为机器语言。其转换过程如图 1-6 所示。



图 1-6 汇编语言程序转换成机器语言程序的过程

高级语言是不针对具体机器的计算机语言,编写程序和阅读程序都比较容易。用高级语言编写的程序,也必须转换成机器语言才能执行,实现这种转换的程序是编译程序和解释程序。

编译程序的功能是把高级语言编写的源程序翻译成目标程序,然后经过链接生成可执行程序,并保存起来。有的高级语言以汇编语言作为中间输出,汇编程序把汇编语言的中间输出变成机器语言(目标程序),链接程序再把目标程序和存放在程序库里的有关信息链接装配在一起,最终产生可执行程序。其转换过程如图 1-7 所示。

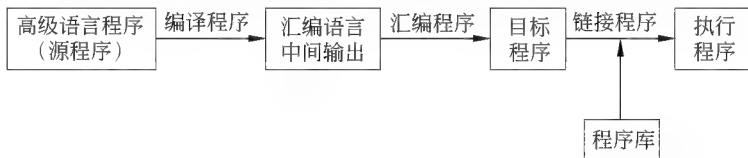


图 1-7 高级语言源程序变成可执行程序的过程

解释程序的功能是对高级语言编写的源程序逐句解释并立即执行,不保留目标程序,不生成可执行程序。

7. 透明性

在计算机中,客观存在的事物或属性从某个角度看不到,就称为“透明”。这与日常生活中的“透明”的含义正好相反。日常生活中的“透明”是要公开,让大家看得到,而计算机

中的“透明”,则是指看不到的意思。

所谓透明实际上就是指那些不属于自己管的部分(不会出现和不需要了解的部分)。通常,在一个计算机系统中,下层机器级的概念性结构和功能特性,对上层机器语言的程序员来说就是透明的。例如,浮点数表示、乘法指令,对高级语言程序员、应用程序员透明,而对汇编语言程序员、机器语言程序员则不透明;再例如,数据总线宽度、微程序对汇编语言程序员、机器语言程序员透明,而对硬件设计者、计算机维修人员则不透明。

1.4 教材习题解答

1. 电子数字计算机和电子模拟计算机的区别在哪里?

解: 电子数字计算机中处理的信息是在时间上离散的数字量,运算的过程是不连续的;电子模拟计算机中处理的信息是连续变化的物理量,运算的过程是连续的。

2. 冯·诺依曼计算机的特点是什么? 其中最主要的一点是什么?

解: 冯·诺依曼计算机的特点如下:

① 计算机(指硬件)应由运算器、存储器、控制器、输入设备和输出设备五大基本部件组成;

② 计算机内部采用二进制来表示指令和数据;

③ 将编好的程序和原始数据事先存入存储器中,然后再启动计算机工作。

第③点是最主要的一点。

3. 计算机的硬件是由哪些部件组成的? 它们各有哪些功能?

解: 计算机的硬件应由运算器、存储器、控制器、输入设备和输出设备五大基本部件组成。它们各自的功能是:

① 输入设备: 把人们编好的程序和原始数据送到计算机中去,并且将它们转换成计算机内部所能识别和接受的信息方式。

② 输出设备: 将计算机的处理结果以人或其他设备所能接受的形式送出计算机。

③ 存储器: 用来存放程序和数据。

④ 运算器: 对信息进行处理和运算。

⑤ 控制器: 按照人们预先确定的操作步骤,控制整个计算机的各部件有条不紊地自动工作。

4. 什么叫总线? 简述单总线结构的特点。

解: 总线是一组能为多个部件服务的公共信息传送线路,它能分时地发送与接收各部件的信息。单总线结构即各大部件都连接在单一的一组总线上,这个总线被称为系统总线。CPU 与主存、CPU 与外设之间可以直接进行信息交换,主存与外设、外设与外设之间也可以直接进行信息交换,而无须经过 CPU 的干预。

5. 简单描述计算机的层次结构,说明各层次的主要特点。

解: 现代计算机系统是一个硬件与软件组成的综合体,可以把它看成是按功能划分的多级层次结构。

第0级为硬件组成的实体。

第1级是微程序级。这级的机器语言是微指令集,程序员用微指令编写的微程序一般是直接由硬件执行的。

第2级是传统机器级。这级的机器语言是该机的指令集,程序员用机器指令编写的程序可以由微程序进行解释。

第3级是操作系统级。从操作系统的基本功能来看,一方面它要直接管理传统机器中的软硬件资源,另一方面它又是传统机器的延伸。

第4级是汇编语言级。这级的机器语言是汇编语言,完成汇编语言翻译的程序叫做汇编程序。

第5级是高级语言级。这级的机器语言就是各种高级语言,通常用编译程序来完成高级语言翻译的工作。

第6级是应用语言级。这一级是为了使计算机满足某种用途而专门设计的,因此这一级语言就是各种面向问题的应用语言。

6. 计算机系统的主要技术指标有哪些?

解: 计算机系统的主要技术指标有:机器字长、数据通路宽度、主存容量和运算速度等。

机器字长是指参与运算的数的基本位数,它是由加法器、寄存器的位数决定的。

数据通路宽度是指数据总线一次所能并行传送信息的位数。

主存容量是指主存储器所能存储的全部信息量。

运算速度与机器的主频、执行什么样的操作、主存本身的速度等许多因素有关。

第 2 章

数据的机器层次表示

2.1 基本内容要求

数据是计算机加工和处理的对象,数据的机器层次表示将直接影响到计算机的结构和性能。本章主要介绍无符号数和带符号数的表示方法、数的定点与浮点表示方法、字符和汉字的编码方法、数据校验码等。熟悉、掌握本章的内容,是学习计算机原理的最基本要求。

学习要求

- ◆ 了解无符号数与带符号数的区别
- ◆ 了解真值和机器数概念
- ◆ 掌握原码、补码、反码表示法和 3 种机器数之间的区别
- ◆ 理解定点数表示法
- ◆ 掌握定点数的表示范围
- ◆ 理解浮点数表示法
- ◆ 掌握浮点数的表示范围
- ◆ 理解规格化浮点数的概念
- ◆ 掌握最小规格化正数与最小正数的区别
- ◆ 理解浮点数阶码的移码表示法
- ◆ 了解 IEEE 754 浮点数标准
- ◆ 理解常见的字符编码方法(ASCII 码)
- ◆ 了解汉字的表示方法
- ◆ 掌握汉字国标码、汉字区位码和汉字机内码的特点和区别
- ◆ 了解二—十进制编码的原理
- ◆ 掌握 8421 码、2421 码和余 3 码的特点
- ◆ 理解奇偶校验码检错的原理

- ◆ 掌握奇偶校验位的形成方法
- ◆ 理解海明校验码检错的原理
- ◆ 了解循环冗余校验码

2.2 误点疑点解惑

1. 真值和机器数的区别

在日常生活中,常用“+”、“-”号加绝对值来表示数值的大小,以这种形式表示的数值在计算机技术中称为“真值”。

由于“+”或“-”号在计算机中是无法识别的,因此需要把数的符号数码化。通常,约定二进制数的最高位为符号位,“0”表示正号,“1”表示负号。这种在计算机中使用的表示数的形式称为机器数,常见的机器数有原码、补码、反码等。

为了能正确地区别出真值和各种机器数,在本套教材中用 X 表示真值, $[X]_{\text{原}}$ 表示原码, $[X]_{\text{补}}$ 表示补码, $[X]_{\text{反}}$ 表示反码。

初学者常容易将真值与原码、补码等机器数混淆,要提醒学生注意:二进制真值前面带有正、负号,虽然正号通常略去不写,但也不要忘记它,尤其是纯整数时。二进制机器数的最高位为 0 表示正数,最高位为 1 表示负数。

2. 模与补码表示法

模是引出补码表示法的一个重要的概念。以时钟为例是说明模的最好方法,如现有时钟正指向 10 点整,但是当前标准时间是 6 点整,为了校准时钟,可顺时针方向拨过 8 个小时(+8),也可逆时针方向拨过 4 个小时(-4),其效果是相同的,如图 2-1 所示。也就是说: $-4 = +8 \pmod{12}$ 。

由此可以得出结论,一个负数可以用一个与它互为补数的正数来代替。将补数的概念用到计算机中,便出现了补码表示法。

计算机本身是一个模数系统,这是因为机器字长是有限的。当运算结果的位数超过机器字长时,向更高位的进位就会被丢失,这就是该计算机的“模”。

对于 $n+1$ 位的定点小数来说,可表示为:

$$X_s . X_1 X_2 \cdots X_n$$

其中, X_s 是符号位,它的位权为 2^0 。符号位向更高位的进位会被丢失,所以定点小数以 $2^1 = 2$ 为模。

对于 $n+1$ 位的定点整数来说,可表示为:

$$X_s X_1 X_2 \cdots X_n$$

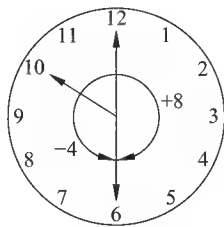


图 2-1 时钟以 12 为模

其中, X_s 仍是符号位, 但它的位权为 2^n 。符号位向更高位的进位会被丢失, 所以定点整数 (字长 $n+1$ 位) 以 2^{n+1} 为模。

3. 原码和补码的区别

在计算机中, 用的最多的机器数是原码和补码。对于正数, 原码和补码没有任何区别, 对于负数, 原码和补码的表示形式完全不同, 且补码要比原码多表示一个最负的数。造成这一现象的直接原因是因为对于真值 0, 原码有两种不同的表示形式, 而补码只有惟一的一种表示形式。

假设有一个字长为 8 位的二进制代码 10000000, 若其为原码, 表示 -0 , 若其为补码, 则不再表示 -0 , 而表示一个绝对值最大的负数。此时最高位的“1”有两个含义, 既代表负号, 又代表这一位的位权。如果这是一个定点整数, 其值等于 $-2^7 = -128$; 如果这是一个定点小数, 其值等于 $-2^0 = -1$ 。

4. 定点数的表示范围

数的表示范围是学生学习中的一个重点和难点, 尤其是浮点数的表示范围更复杂, 但是切不可因此而把注意力仅放在浮点数的学习上。应当明确浮点数是由定点小数和定点整数共同组成的, 所以首先要搞清楚定点小数和定点整数的表示范围, 否则对浮点数表示范围的学习就无法下手。

建议在这部分内容的教学过程中不要仅仅告诉学生相关几个点 (最大正数——数轴上最右边的点、最小正数——数轴上正数区最接近于零的点、绝对值最大的负数——数轴上最左边的点) 的值, 而让学生去死记硬背。一定要给学生讲清楚这几个点在原码表示和补码表示时的代码形式, 并介绍如何根据二进制数的位权来得到对应的数值, 否则学生没有举一反三的能力, 题目稍微有些变化就不会做了。

下面首先看定点小数的表示范围, 图 2-2 给出了最大正数和最小正数的表示形式, 由于同一正数的原码和补码表示形式完全相同, 所以这里就不再区分原码和补码了。注意: 在图 2-2 中每一位二进制数上都标出了这一位的位权, 小数点在符号位和最高有效数位之间。

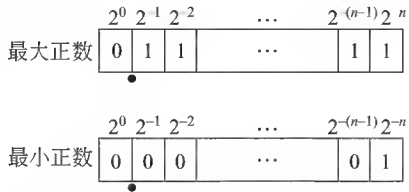


图 2-2 定点小数表示的最大正小数和最小正小数

从图 2-2 中可以看出最大正数的数值位部分全部为 1, 如果要写出它的真值来, 则应

该是：

$$\text{最大正数} = 2^{-1} + 2^{-2} + \dots + 2^{-n}$$

这个值似乎显得太复杂了一点,我们不妨可以用一个更简单的方法来表述它。即

$$\text{最大正数} = 1 - 2^{-n}$$

这个简化了的值是根据如下关系得到的：

$$\begin{array}{rcl} 100\dots\dots 00 & & 2^0 = 1 \\ -000\dots\dots 01 & & 2^{-n} \\ \hline 011\dots\dots 11 & & 1 - 2^{-n} \end{array}$$

同理,从图 2-2 可以看到最小正数的数值位最低位为 1,其真值当然等于 2^{-n} 。

图 2-3 中分别给出了原码和补码绝对值最大负数的表示形式。显然,原码与补码所表示的绝对值最大的负数是有区别的。

	2^0	2^{-1}	2^{-2}	...	$2^{-(n-1)}$	2^{-n}
原码表示的绝对值最大的负数	1	1	1	...	1	1
	2^0	2^{-1}	2^{-2}	...	$2^{-(n-1)}$	2^{-n}
补码表示的绝对值最大的负数	1	0	0	...	0	0

图 2-3 原码和补码表示的绝对值最大负数(定点小数)

因为在原码表示时,正数和负数范围是对称的,所以绝对值最大的负数等于最大正数值加上“一”号,其真值等于 $-(1 - 2^{-n})$ 。

从图 2-3 中很容易看出补码表示的绝对值最大的负数值等于 -1。

接下来看定点整数的表示范围,图 2-4 给出了最大正数和最小正数的表示形式,在图 2-4 中每一位二进制数上也标出了这一位的位权,此时小数点在最低有效数位之后。

	2^n	2^{n-1}	2^{n-2}	...	2^1	2^0
最大正数	0	1	1	...	1	1
	2^n	2^{n-1}	2^{n-2}	...	2^1	2^0
最小正数	0	0	0	...	0	1

图 2-4 定点整数表示的最大正数和最小正数

根据前述定点小数的结果,很容易推出：

$$\text{最大正数} = 2^n - 1$$

$$\text{最小正数} = 1$$

图 2-5 中分别给出了原码和补码绝对值最大负数的表示形式。

	2^n	2^{n-1}	2^{n-2}	...	2^1	2^0
原码表示的绝对值最大的负数	1	1	1	...	1	1
	2^n	2^{n-1}	2^{n-2}	...	2^1	2^0
补码表示的绝对值最大的负数	1	0	0	...	0	0

图 2-5 原码和补码表示的绝对值最大负数(定点整数)

不难推出：

原码表示绝对值最大负数 $=-(2^n-1)$ 。

补码表示绝对值最大负数 $=-2^n$ 。

5. 浮点数的表示范围

$$\text{浮点数 } N = M \times r^E$$

式中： r 是浮点数阶码的底，也称为尾数基数，通常 $r=2$ 。 E (阶码部分)和 M (尾数部分)都是带符号的定点数，在大多数计算机中，尾数为纯小数，常用原码或补码表示；阶码为纯整数，常用移码或补码表示。

一般的浮点数格式如图 2-6 所示，这个浮点数的总长度为 $k+n+2$ 位。

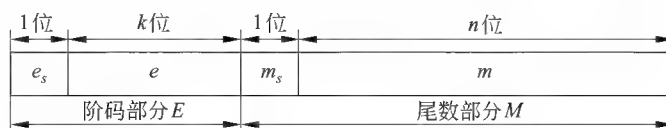


图 2-6 浮点数的一般格式

假设浮点数的尾数和阶码均用补码表示。有了前述的定点小数和定点整数的基础，推出浮点数的表示范围应该是不困难的。图 2-7 中分别给出了浮点数的最大正数、最小正数和绝对值最大负数的表示形式。

显然，浮点数的最大正数应当是阶码部分和尾数部分都为最大正数，其值等于：

$$\text{最大正数} = (1-2^{-n}) \times 2^{2^k-1}$$

浮点数的最小正数应当是尾数部分为最小正数，阶码部分为绝对值最大的负数，其值等于：

$$\text{最小正数} = 2^{-n} \times 2^{-2^k}$$

浮点数的绝对值最大负数应当是尾数部分为绝对值最大的负数，阶码部分为最大正数，其值等于：

$$\text{绝对值最大负数} = -1 \times 2^{2^k-1}$$

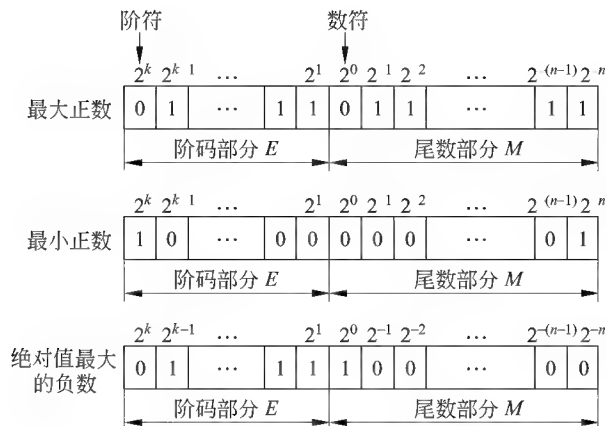


图 2-7 浮点数中几个关键点的代码表示形式

6. 浮点数的规格化

为了提高运算的精度,通常规定参加运算的浮点数必须是规格化形式的。那么什么是规格化的浮点数呢?规格化浮点数的特点是尾数的最高数位必须是一个有效值。假设尾数的基数 $r=2$ 。

正数的情况比较简单,无论尾数用原码还是补码表示,其规格化形式均为: $0.1 \times \dots \times$ 。

负数的情况比较复杂,若尾数用原码表示,则规格化形式为: $1.1 \times \dots \times$;若尾数用补码表示,则规格化形式为: $1.0 \times \dots \times$ 。原码的结果很容易接受,而补码的这一结果往往难以理解。此时应该说,尾数的最高数位必须是一个有效值,并不是指机器数(原码和补码)的最高数位必须是 1,而是指真值的最高数位必须是 1。

设浮点数的尾数和阶码均用补码表示,规格化的最小正数和规格化的绝对值最小负数的表示形式如图 2-8 所示。

很明显,规格化的最小正数的真值为:

$$\text{规格化的最小正数} = 2^{-1} \times 2^{-2^k}$$

规格化的绝对值最小负数的真值为:

$$\text{规格化的绝对值最小负数} = -(2^{-1} + 2^{-n}) \times 2^{-2^k}$$

7. 移码偏置值的选择

在许多通用计算机中,常采用移码表示浮点数的阶码。用移码来表示阶码有两个优点:

- 便于比较浮点数的大小。
- 简化机器中的判零电路。

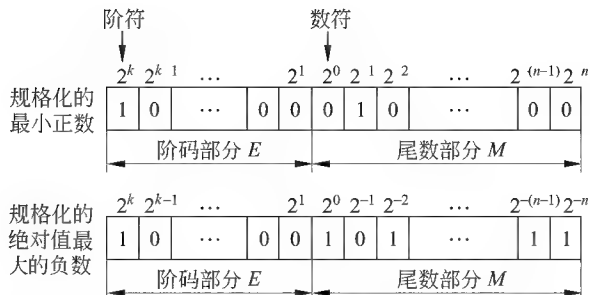


图 2.8 规格化的最小正数和规格化的绝对值最小负数的表示形式

移码需要在真值 X 基础上加一个常数,这个常数被称为偏置值。

$$[X]_{\text{移}} = \text{偏置值} + X$$

关于偏置值的选择是一个需要考虑的问题,在主教材 2.2.3 节和 2.2.6 节中两种浮点数使用的两个偏置值是不同的。那么如何来选择偏置值呢? 移码可以被视为无符号数,对于字长为 8 位的移码,总共有 2^8 个无符号数:

00000000	0	
00000001	1	
...		
01111111	$2^7 - 1$	} 居于中间的两个数
10000000	2^7	
...		
11111110	$2^8 - 2$	
11111111	$2^8 - 1$	

2^8 个无符号数对应于 2^8 个阶码的真值,为使阶码真值的正数和负数分布尽可能均匀,可以选择居于中间的两个数中的任何一个作为偏置值。

在 IEEE 754 标准中,明确给出了移码的偏置值,其中短浮点数的偏置值为 $2^7 - 1$,长浮点数的偏置值为 $2^{10} - 1$ 。但对于没有明确指出偏置值的移码,应选择偏置值 2^n (设字长为 $n+1$ 位)。

对于偏置值为 2^n 的移码来说,同一数值的移码和补码除最高位相反之外,其他各位相同。由此可见,对应于同一个真值,浮点数的阶码用移码表示和用补码表示的形式是不相同的。

8. 定点数与浮点数的比较

对于某一种数的表示方法,主要关心它的两项指标:一项是表示范围,即这种方法能表示数值的大小(正负两个方向);另一项是精度,也称分辨率,即精细的程度。这就好比

一把尺子一次测量的范围由其长度决定,而尺子的精度由它的最小刻度决定。在数轴上非零的最小正数这个典型值就是分辨率。

例 2-1 比较字长为 32 位的定点整数和浮点数的表示范围和精度。

解: 若定点整数 32 位,补码表示,则表示范围为 $-2^{31} \sim (2^{31}-1)$,分辨率为 1。

若浮点数 32 位,其中阶码部分 8 位,含 1 位阶符,补码表示,以 2 为底;尾数部分 24 位,含 1 位数符,补码表示,规格化。则表示范围为 $-2^{127} \sim 2^{127} \times (1-2^{-23})$,最高分辨率为 2^{-129} 。

显然,浮点数的表示范围比定点数大得多。这里会让人们产生一种误解,为什么浮点数的表示范围比定点数大得多,而且分辨精度也高得多呢? 不是说,浮点数扩大了数的表示范围,是以降低精度为代价的吗? 其实,浮点数的分辨率 2^{-129} 只是该浮点格式下的最高分辨率,它对应于阶码为绝对值最大负数时,当阶码值增大时,分辨率将随之降低(值变大),而阶码值减少时,分辨率随之提高(值变小),这是因为浮点数与定点整数在数轴上的分布存在着很大的不同。

整数在数轴上是均匀分布的,也就是说连续两个数据之间的差都为 1,所以定点整数的分辨率为 1。

浮点数在数轴上分布则是不均匀的,越是靠近零点,数越密集,越远离零点,数越稀疏。原因是:浮点数的位数确定了,则所能表示的数据个数就确定了。对于基数为 2 的情况,阶码绝对值为 $n+1$ 的浮点数值覆盖区域比阶码绝对值为 n 的浮点数值覆盖区域大一倍,但两者在各自区域中所能表示的数据个数是相同的,因此,阶码绝对值为 $n+1$ 的浮点数覆盖区域内的数据密度比阶码绝对值为 n 的浮点数值浮点区域小一倍。浮点数的数据密度分布如图 2-9 所示。

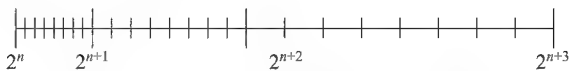


图 2-9 浮点数的数据密度分布示意

9.3 种汉字编码的区别

汉字国标码是指 GB2312 80 标准。GB2312 80 标准共包括 6763 个汉字,按其使用频率,分为一级汉字(3755 个)和二级汉字(3008 个)。在 GB2312 80 标准中,所有符号按区位编排,共设 94 区,每区含 94 个汉字和符号。一级汉字按拼音顺序排列,占据 16~55 区;二级汉字按部首顺序排列,占据 56~87 区;前 15 区用来编排西文字母、数字、图形符号,以及用户自行定义的专用符号;目前,10~15 区空着。国标码(GB2312 80)编排如表 2-1 所示。

表 2-1 GB2312—80 汉字编码

第 2 字节		位号	1	...	94
第 1 字节		国标码	21H	...	7EH
区号	国标码				
1	21H		字母、数字、图符		
⋮	⋮				
7	27H				
⋮	⋮		一级汉字(3008 个)		
16	30H				
⋮	⋮				
55	57H		二级汉字(3008 个)		
56	58H				
⋮	⋮				
87	77H				
88	78H				
⋮	⋮				
94	7EH				

汉字国标码用两个字节的十六进制数表示,每个国标码都有一个惟一对应的十进制区号和位号。

汉字机内码是汉字在计算机内部的编码,它也是两字节长的代码,机内码是在相应国标码的每个字节最高位上加“1”,即:

$$\text{汉字机内码} = \text{汉字国标码} + 8080\text{H}$$

汉字区位码是一种输入码,区位码长 4 位,前 2 位表示区号,后 2 位表示位号,汉字的区号和位号均用十进制数表示。区位码与国标码有简单的对应关系:

$$\text{汉字国标码} = \text{汉字区位码(十六进制)} + 2020\text{H}$$

需要特别提醒学生注意的是,汉字的区位码是用十进制数表示的,通常记作“区号一位号”。在 3 种汉字编码的转换时,千万不要忘记先将十进制的区位码变成十六进制之后,再利用上述关系式进行转换。

10. 十进制数的 BCD 编码

虽然常见的 BCD 码只有几种,如 8421 码、2421 码和余 3 码等。但实际上 BCD 码有很多种,这是因为 4 位二进制数可以表示 16 种不同的状态,现只需要使用其中的 10 种状态来表示 0~9 这 10 个数码,在 16 种不同的状态中取任意的 10 种状态的方法有很多种。所以说 BCD 码是有冗余状态的编码。

BCD 码用 4 位二进制数来表示 1 位十进制数,如十进制数 3609 可以分别表示为:

$$\begin{aligned}(3609)_{10} &= (0011\ 0110\ 0000\ 1001)_{8421\text{码}} \\ &= (0011\ 1100\ 0000\ 1111)_{2421\text{码}} \\ &= (0110\ 1001\ 0011\ 1100)_{\text{余3码}}\end{aligned}$$

应当注意的是,有些学生可能会把 8421 码与 BCD 码混为一谈。产生这种误解的主要原因在于一些“微型计算机原理”的教材中常将 BCD 码当作 8421 码,由于“微型计算机原理”总是针对某种具体机型的,在 80x86 中使用的 BCD 码恰恰是 8421 码,所以在“微型计算机原理”中将 BCD 码当作 8421 码不能算作错误,但毕竟这是不准确的。“计算机组成原理”是不拘泥于某一种具体机型的,严格地说,8421 码只是 BCD 码中的一种形式而已,不能说 BCD 码就是 8421 码。

还应当注意的是,在 8421 码中 0~9 这 10 个数码的表示形式与用二进制表示的形式一样,但这是两个完全不同的概念,不能混淆。例如,一个两位的十进制数 39,它可以表示为:

$$(0011\ 1001)_{8421\text{码}} \text{ 或 } 1001111\text{B}$$

这两者是完全不同的。

11. 奇偶校验位的形成和奇偶校验码的检测

奇偶校验码是由若干位有效信息位,再加上一个二进制位(校验位)组成的。校验位的取值(0 或 1)将使整个校验码中“1”的个数为奇数或偶数,有两种可供选择的校验规律:

奇校验 —— 整个校验码中“1”的个数为奇数。

偶校验 —— 整个校验码中“1”的个数为偶数。

首先要给学生讲清楚奇偶校验位和奇偶校验码的区别,奇偶校验位只有 1 位,而奇偶校验码共 $n+1$,不仅包括奇偶校验位,还包括所有的 n 位有效信息位。然后再介绍校验位的形成和校验码的检测方法。主存读写过程中的奇偶检验示意如图 2-10 所示。

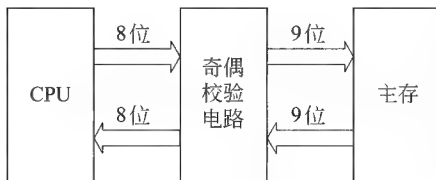


图 2-10 主存读写过程中的奇偶检验示意

假设 CPU 准备写入主存某单元的数据为 01010101,若采用奇校验,经过奇偶校验电路,形成奇偶校验位,实际写入主存的 9 位校验码是 101010101(最高位是检验位)。从主存单元读出的 9 位信息首先送入奇偶校验电路进行检测,若 9 位信息中“1”的个数为奇数个,表示读出信息正确,将校验位去掉之后的 8 位数据送 CPU;若 9 位信息中“1”的个数为偶数个,表示读出信息不正确,向 CPU 发出奇偶校验出错的中断请求信号。

奇偶校验位的形成及校验电路如图 2-11 所示。图 2-11 的虚线框中为检验位形成电路,7 个异或门实际上是在数 8 个数据位 $D_7 \sim D_0$ 中“1”的个数。

$$\text{偶形成} = D_7 \oplus D_6 \oplus D_5 \oplus D_4 \oplus D_3 \oplus D_2 \oplus D_1 \oplus D_0$$

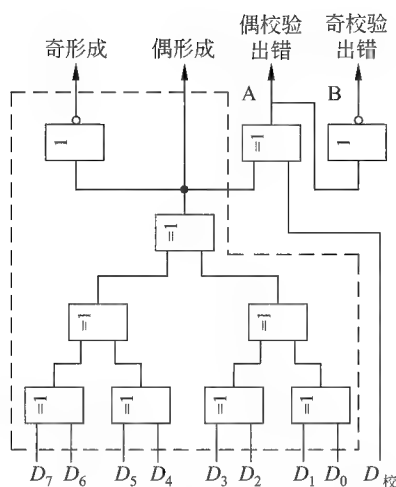


图 2 11 奇偶校验位的形成及校验电路

$$\text{奇形成} = D_7 \oplus D_6 \oplus D_5 \oplus D_4 \oplus D_3 \oplus D_2 \oplus D_1 \oplus D_0$$

形成的校验位和数据位 $D_7 \sim D_0$ 一起写入主存。

读出时,9 位代码同时送入奇偶校验电路检测,8 个异或门实际上是在数 9 位代码中“1”的个数。

$$\text{偶校验出错} = D_{\text{校}} \oplus D_7 \oplus D_6 \oplus D_5 \oplus D_4 \oplus D_3 \oplus D_2 \oplus D_1 \oplus D_0$$

$$\text{奇校验出错} = D_{\text{校}} \oplus D_7 \oplus D_6 \oplus D_5 \oplus D_4 \oplus D_3 \oplus D_2 \oplus D_1 \oplus D_0$$

若 9 位代码中“1”的个数符合奇偶校验码的要求,则检验出错位=0,反之检验出错位=1。

2.3 相关知识介绍

1. 补码 $[X]_{\text{补}}$ 与真值 X 的转换

$$X \geq 0, [X]_{\text{补}} = X$$

$$X < 0, [X]_{\text{补}} = M + X$$

例 2-2 以定点小数为例,设 $[X]_{\text{补}} = X_s . X_1 X_2 \cdots X_n$

求证 1:

$$X = -X_s + \sum_{i=1}^n X_i 2^{-i}$$

证明:

当 $X \geq 0$ 时, $X_s = 0$,

$$[X]_{\text{补}} = 0.X_1 X_2 \cdots X_n = \sum_{i=1}^n X_i 2^{-i} = X$$

当 $X < 0$ 时, $X_s = 1$,

$$[X]_{\text{补}} = 1.X_1 X_2 \cdots X_n = 2 + X$$

$$\begin{aligned} \text{所以} \quad X &= 1.X_1 X_2 \cdots X_n - 2 = -1 + 0.X_1 X_2 \cdots X_n \\ &= -1 + \sum_{i=1}^n X_i 2^{-i} \end{aligned}$$

综合上述情况,可以得出:

$$X = -X_s + \sum_{i=1}^n X_i 2^{-i}$$

求证 2:

$$[X]_{\text{补}} = 2X_s + X, \text{其中 } X_s = \begin{cases} 0 & 0 \leq X < 1 \\ 1 & -1 \leq X < 0 \end{cases}$$

证明:

当 $0 \leq X < 1$ 时,则

$$0 \leq [X]_{\text{补}} = X < 1$$

因为正数补码等于正数本身,所以

$$0 \leq X_s.X_1 X_2 \cdots X_n < 1, X_s = 0$$

当 $-1 \leq X < 0$ 时,根据补码定义有:

$$1 \leq [X]_{\text{补}} = 2 + X < 2 (\text{mod } 2)$$

$$1 \leq X_s.X_1 X_2 \cdots X_n < 2, X_s = 1$$

若 $0 \leq X < 1, X_s = 0$,则 $[X]_{\text{补}} = 2X_s + X = X$

若 $-1 \leq X < 0, X_s = 1$,则 $[X]_{\text{补}} = 2X_s + X = 2 + X$

$$\text{所以有: } [X]_{\text{补}} = 2X_s + X, \text{其中 } X_s = \begin{cases} 0 & 0 \leq X < 1 \\ 1 & -1 \leq X < 0 \end{cases}$$

2. 原码 $[X]_{\text{原}}$ 与补码 $[X]_{\text{补}}$ 的转换

下面分两种情况讨论。

(1) 正数的 $[X]_{\text{补}}$ 与 $[X]_{\text{原}}$ 的关系

当 $X \geq 0$ 时,

因为

$$[X]_{\text{原}} = X, [X]_{\text{补}} = X$$

所以

$$[X]_{\text{补}} = [X]_{\text{原}}$$

即正数的补码等于它的原码。

(2) 负数的 $[X]_{\text{补}}$ 与 $[X]_{\text{原}}$ 的关系

当 $X < 0$ 时, $[X]_{\text{补}}$ 等于把 $[X]_{\text{原}}$ 除去符号位外的各位求反后再加“1”。

例 2-3 以字长为 $n+1$ 位的定点整数为例,证明负数的 $[X]_{\text{补}}$ 与 $[X]_{\text{原}}$ 的关系。

已知: $[X]_{\text{原}} = 1, X_1 X_2 \cdots X_n$ 。求证: $[X]_{\text{补}} = 1, \overline{X_1 X_2 \cdots X_n} + 1$

证明:

因为

$$[X]_{\text{原}} = 2^n - X, [X]_{\text{补}} = 2^{n+1} + X$$

$$[X]_{\text{原}} + [X]_{\text{补}} = 2^{n+1} + 2^n$$

所以

$$\begin{aligned} [X]_{\text{补}} &= 2^{n+1} + 2^n - [X]_{\text{原}} \\ &= 2^{n+1} + 2^n - 1, X_1 X_2 \cdots X_n \\ &= 2^{n+1} + 2^n - 2^n - X_1 X_2 \cdots X_n \\ &= 2^{n+1} - X_1 X_2 \cdots X_n \\ &= 2^n + 2^n - X_1 X_2 \cdots X_n \\ &= 2^n + (2^n - 1) + 1 - X_1 X_2 \cdots X_n \\ &= 2^n + (2^n - 1) - X_1 X_2 \cdots X_n + 1 \end{aligned}$$

注意: 因为 $2^n - 1 = 11 \cdots 1$, 共计 n 个1, 而 $11 \cdots 1 - X_1 X_2 \cdots X_n$ 就是对每位数码 X_i 求反。

所以

$$\begin{aligned} [X]_{\text{补}} &= 2^n + X_1 X_2 \cdots X_n + 1 \\ &= 1, X_1 X_2 \cdots X_n + 1 \end{aligned}$$

同理已知: $[X]_{\text{补}} = 1, X_1 X_2 \cdots X_n$, 则

$$\begin{aligned} [X]_{\text{原}} &= 2^{n+1} + 2^n - [X]_{\text{补}} \\ &= 1, \overline{X_1 X_2 \cdots X_n} + 1 \end{aligned}$$

即 $[X]_{\text{原}} = [[X]_{\text{补}}]_{\text{补}}$

3. 浮点数的表数范围

由于机器字长有限,浮点数只能表示出数轴上分散于正、负两个区间中的部分离散值,浮点数的表示范围在数轴上的表示如图 2-12 所示,图中阴影部分是数的表示范围,规格化浮点数的表示范围小于非规格化浮点数的表示范围。

在浮点数中,当数据的绝对值太大,以至于大于阶码所能表示的数(阶码上溢)时,称为浮点数的上溢。而当数据的绝对值太小,以至于小于阶码所能表示的数(阶码下溢)时,则称为浮点数的下溢。

只要浮点数的尾数为全0,不论阶码为何值,一般都当作机器零处理。为了保证浮点数0表示形式的惟一性,此时应把阶码置成最小值(绝对值最大的负数)。

4. 浮点数尾数基数的选择

浮点数由阶码和尾数两部分组成,两者都有各自的基数。但在实际应用中,为了简

单,阶码的基数都为2,而尾数的基数 r 则可以为2、4、8或16。因此,浮点数的基数选择问题实际上仅仅是指尾数的基数选择。

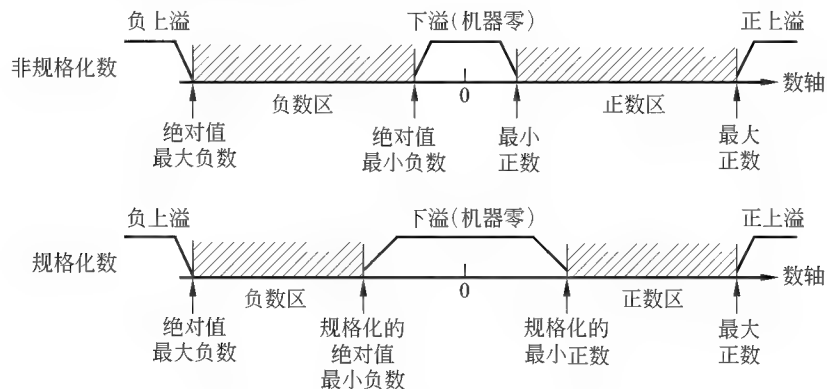


图 2 12 浮点数的表示范围

当尾数的基数为8或16时,浮点数表示成

$$N = M \times 8^E \text{ 或 } N = M \times 16^E$$

此时阶码 E 和尾数 M 仍用二进制表示,其运算规则也基本与基数 $r=2$ 时相同,只是在执行对阶和规格化操作时,是以基数 r 为尺度进行移位的,尾数左移(或右移) $\log_2 r$ 位,阶码减(或加)1。

当浮点数的总位数给定的情况下,所选择的基数越大,所表示的数的范围就越大。

假定某浮点数字长32位,阶码部分(阶符和阶码数值位)共8位,尾数部分(数符与尾数数值位)共24位,均用补码表示。

若 $r=2$,浮点数的表示范围为:

$$-1 \times 2^{2^7-1} \leq X \leq (1 - 2^{-23}) \times 2^{2^7-1}$$

若 $r=16$,浮点数的表示范围为:

$$-1 \times 16^{2^7-1} \leq X \leq (1 - 2^{-23}) \times 16^{2^7-1}$$

当尾数的基数为8或16时,判断尾数是否为规格化数时,应使尾数的数值位的最高3位或4位中至少有1位与符号位不同(补码)。规格化浮点数的尾数表示范围分别为:

$$-1 \leq M < -\frac{1}{8} \text{ 或 } \frac{1}{8} \leq M < 1$$

$$-1 \leq M < -\frac{1}{16} \text{ 或 } \frac{1}{16} \leq M < 1$$

若 $r=8$,则 $|M| \geq \frac{1}{8}$,即 $|M| \geq 0.001 \times \dots \times$ 。

若 $r=16$, 则 $|M| \geq \frac{1}{16}$, 即 $|M| \geq 0.0001 \times \dots \times$ 。

各种浮点数计算机中,为什么要采用不同的尾数的基数呢?下面以一个具体例子来说明。

某机中有这样一个浮点数:

0,010,0,10110111

假设尾数的基数不同,于是:

若 $r=2$, 则 $N=0.10110111 \times 2^2 = (10.110111)_2 \approx (2.75)_{10}$

若 $r=8$, 则 $N=0.10110111 \times 8^2 = (101101.11)_2 = (45.75)_{10}$

若 $r=16$, 则 $N=0.10110111 \times 16^2 = (10110111)_2 = (183)_{10}$

可以看出,阶码相同,尾数也相同的一个浮点数,只因为它们选择了不同的尾数基数,所表示的浮点数值各不相同,而且尾数的基数越大,所表示的浮点数值就越大。这与加长阶码的长度可取得相同的效果——增大了浮点数的表示范围,这就是各种浮点计算机中采用不同的尾数基数的原因。例如 PDP-11 和 IBM 370 的短浮点数具有同样的格式,但前者 $r=2$, 后者 $r=16$ 。所以 IBM 370 的短浮点数比 PDP-11 的短浮点数的表示范围要大,但相对误差也较大。

5. IEEE 754 标准的浮点数

计算机中的浮点数一般都是用二进制表示的。如果在不同的计算机中,浮点数采用不同的基数、尾数和阶码的长度,则浮点数表示有较大的差别,这样不利于软件在不同的机器之间的移植。从 20 世纪 70 年代末开始,IEEE 就成立了一个专门的委员会负责对浮点数进行标准化。IEEE 754 标准是 1985 年由 IEEE 提出的一个从系统结构角度支持浮点数表示的标准,当今流行的计算机几乎都采用这一标准。

在 IEEE 754 标准中阶码用移码表示,尾数用原码表示,隐含的基数为 2。表 2-2 总结了短浮点数(32 位)和长浮点数(64 位)格式的有关参数。

表 2 2 IEEE 754 格式参数

参 数	短浮点数(32 位)	长浮点数(64 位)
总位数	32	64
阶码位数	8	11
阶码偏置值	127	1023
阶码最大值	127	1023
阶码最小值	-126	-1022
数的范围(基数为 10)	$10^{-38}, 10^{+38}$	$10^{-308}, 10^{+308}$

续表		
参 数	短浮点数(32 位)	长浮点数(64 位)
尾数位数	23	52
阶码数目	254	2046
尾数数目	2^{23}	2^{52}

以短浮点数为例,阶码最大值为 127,最小值为-126,则阶码移码的表示范围为 1~254。这是因为阶码为全 0 和全 1 这两种极端阶码值用于定义特殊数值:机器零和无穷大。

非 0 规格化数的尾数的最高有效位一定为 1。IEEE 754 标准规定规格化浮点数在小数点的左边有一隐含位(作为二进制整数的个位数)。由于该位为 1,不需要存储,在运算时,自动加上该位参加运算,因此尾数实际上是 24 位。此时规格化浮点数的尾数为 1.f (f 为尾数,1 为隐含位),所表示的规格化浮点数为 $\pm 2^{E-127} \times (1.f)$ 。

6. 汉字的字形码

在计算机中的汉字机内码不可能直接在屏幕上显示和在打印机上输出,必须把它转换成对应的汉字字形码。汉字字形码是指确定一个汉字字形点阵的代码,又称作汉字字模码或汉字输出码。在一个汉字点阵中,凡笔画所到之处,记为“1”,否则记为“0”。

根据对汉字质量的不同要求,可有 16×16、24×24、32×32 或 48×48 的点阵结构。显然点阵越大,输出汉字的质量越高,每个汉字所占用的字节数也越高,如表 2-3 所示。

表 2 3 汉字点阵分类

字型	点阵(行×列)	字节数/B	特征
简易型	16×16	32	显示字体骨架
普及型	24×24	72	有笔锋,可分字体
提高型	32×32	128	笔锋清晰,字体齐全
精密型	48×48	288	能表示复杂字型

汉字字模库又称作汉字库,用来存储汉字点阵字模信息,常由 ROM 或磁盘组成。存储器为 ROM 的汉字库称作硬字库,硬字库不占用用户可用主存空间;存储器为磁盘的汉字库称作软字库,使用时需把汉字点阵存入主存。汉字库的信息量很大,一个 16×16 点阵的汉字库,至少需要 256KB,而 24×24 点阵的汉字库需 576KB,32×32 点阵的汉字库则需 1MB。

汉字点阵是以字节为单位存储的,考虑到存储器的限制以及设计上的方便,一般的汉字系统所使用的汉字点阵为 16×16 点阵和 24×24 点阵,但这两种点阵汉字的存储方式不同。

16×16 点阵字模的存储方式是按行存储,一个字节存放一个行点阵码,如图 2-13

所示。

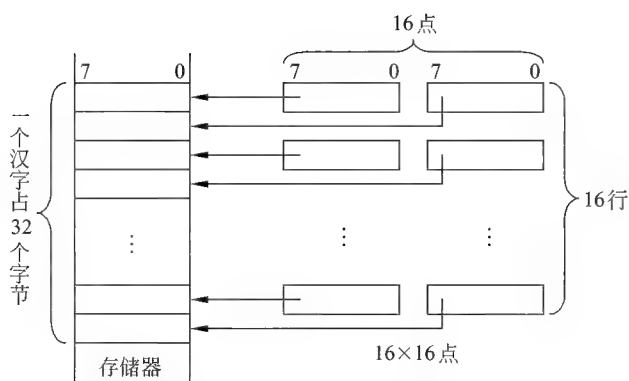


图 2 13 16×16 点阵字模的存储方式

24×24 点阵字模的存储方式是按列存储,一个字节存放一个列点阵码,如图 2-14 所示。

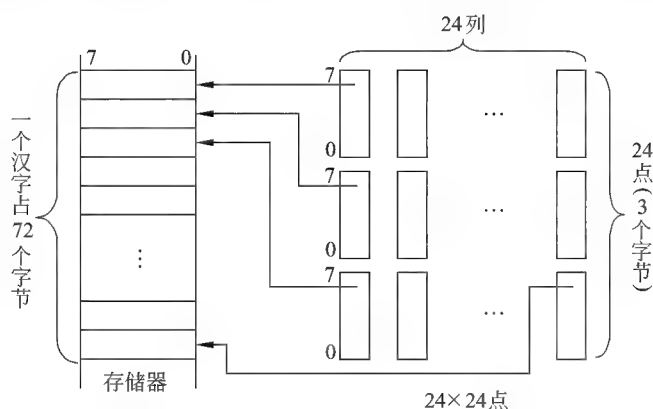


图 2 14 24×24 点阵字模的存储方式

7. 校验码的码距

任何一种编码都由许多码字构成,任何两个相邻码字之间会有 n 位代码不同,这就被称作是它们之间的距离,这些 n 值中,最小的值就是该种编码的码距。

例如,BCD 码共包含 10 个码字,以 8421 码为例,它们的顺序为 0000,0001,0010,0011,...,1000,1001。任意两个相邻码字之间的距离各不相同,如 0000 与 0001,0010 与 0011 之间的距离为“1”,0111 与 1000 之间的距离为“4”,所以 8421 码的码距 $L=1$ 。这种编码没有检错能力,因为当某一个合法码字中有一位出错,就变成为另一个合法码字了。

具有检、纠错能力的数据校验码的实现原理是:在编码中,除去合法的码字外,再加进

一些非法的码字,当某个合法码字出现错误时,就变成某个非法码字。合理地安排非法码字的数量和编码规则,就能达到纠错的目的。例如,加上奇偶校验位的 8421 码,以偶校验为例,10 个码字依次为:00000,10001,10010,00011,10100,00101,00110,10111,11000,01001,任意两个相邻码字之间的距离均大于或等于 2,所以带奇偶校验的 8421 码的码距 $L=2$ 。如果上述码字中有一位出错,会造成结果变成一个非法码,即代码中“1”的个数不是偶数个。

在纠错理论中,有一个重要公式:

$$L-1 = C + D \quad \text{且 } D \geq C$$

其中: L ——编码的码距;

C ——可以纠错的位数;

D ——可以检错的位数。

从上式可以看出,编码的纠错、检错能力与码距密切相关。对于 $L \geq 2$ 的数据校验码,开始具有检错的能力。码距越大,检错和纠错能力就越强,而且检错能力应大于或等于纠错能力。

8. 海明编码

主教材在讨论能检测和自动校正一位错并能发现两位错的海明码时,提到校验位的位数 K 和信息位的位数 N 应满足下列关系: $2^{K-1} \geq N + K + 1$ 。如果不考虑码距均匀的问题,只要满足 $2^K \geq N + K + 1$ 就可以了。因此能纠正一位错的海明码的最少校验位数如表 2-4 所示。

表 2 4 能纠正一位错的海明码的最少校验位数

字长	校验位	总长	校验位与字长之比 /%
8	4	12	50
16	5	21	31
32	6	38	19
64	7	71	11
128	8	136	6
256	9	265	4
512	10	522	2

例如,对于 16 位长的数据,需加入 5 位校验位,海明码一共有 21 位。海明码位号与检验位、数据位的对应关系如表 2-5 所示。

表 2 5 海明码位号与检验位、数据位的对应关系

海明码位号	H ₂₁	H ₂₀	H ₁₉	H ₁₈	H ₁₇	H ₁₆	H ₁₅	H ₁₄	H ₁₃	H ₁₂	H ₁₁
数据/检验位	D ₁₆	D ₁₅	D ₁₄	D ₁₃	D ₁₂	P ₅	D ₁₁	D ₁₀	D ₉	D ₈	D ₇
海明码位号	H ₁₀	H ₉	H ₈	H ₇	H ₆	H ₅	H ₄	H ₃	H ₂	H ₁	
数据/检验位	D ₆	D ₅	P ₄	D ₄	D ₃	D ₂	P ₃	D ₁	P ₂	P ₁	

从表 2-5 中可以看出,海明码的第 1、2、4、8 和 16 位为校验位,其他位都是数据位。每位校验位负责校验的海明码位分别为:

第 1 位:1,3,5,7,9,11,13,15,17,19,21。

第 2 位:2,3,6,7,10,11,14,15,18,19。

第 4 位:4,5,6,7,12,13,14,15,20,21。

第 8 位:8,9,10,11,12,13,14,15。

第 16 位:16,17,18,19,20,21。

即第 b 位由第 $b_1 b_2 \dots b_j$ 位一起来校验,其中 $b_1 + b_2 + \dots + b_j = b$ 。例如,海明码的第 5 位由第 4 位和第 1 位校验,因为 $5 = 4 + 1$;海明码的第 10 位由第 8 位和第 2 位校验,因为 $10 = 8 + 2$;海明码的第 21 位由第 16 位、第 4 位和第 1 位校验,因为 $21 = 16 + 4 + 1$ 。

9. 循环冗余校验码的模 2 运算

循环冗余校验码在编码、译码时采用的是模 2 运算,即二进制运算时不考虑进位和借位。

(1) 模 2 加减

按位加或减,用异或逻辑实现,有下列规则:

$$\begin{array}{ll} 0 \pm 0 = 0 & 0 \pm 1 = 1 \\ 1 \pm 0 = 1 & 1 \pm 1 = 0 \end{array}$$

例如:

$$\begin{array}{r} 1010 \\ + 0110 \\ \hline 1100 \end{array} \quad \begin{array}{r} 1010 \\ - 0110 \\ \hline 1100 \end{array} \quad \begin{array}{r} 1010 \\ + 1010 \\ \hline 0000 \end{array}$$

由以上例子可见,模 2 加与模 2 减等同,相同两数的模 2 加减结果为 0。

(2) 模 2 乘法

按模 2 加规则计算部分积之和,不进位。

例如:

$$\begin{array}{r} 1010 \\ \times 101 \\ \hline 1010 \\ 0000 \\ 1010 \\ \hline 100010 \end{array} \quad \left. \begin{array}{l} 1010 \\ 0000 \\ 1010 \end{array} \right\} \text{部分积}$$

(3) 模 2 除法

按模 2 减规则求部分余数,不借位。每求一位商应使部分余数减少一位;当部分余数的首位为 1 时,商取 1;当部分余数的首位为 0 时,商取 0;当部分余数的位数小于除数的

位数时,该余数就是最后余数。

例如:

$$\begin{array}{r}
 101 \text{ ——商} \\
 101 \overline{) 10000} \\
 \underline{101} \\
 010 \\
 \underline{000} \\
 100 \\
 \underline{101} \\
 01 \text{ ——余数}
 \end{array}$$

10. 循环冗余校验码的编码和校验过程

循环冗余校验码由信息位和校验位两部分组成,若信息位为 N 位,校验位为 K 位,则该校验码被称为 $(N+K, N)$ 码。

循环冗余校验码的编码规则:

- ① 把待编码的 N 位有效信息表示为多项式 $M(X)$ 。
- ② 把 $M(X)$ 左移 K 位,以便拼装 K 位余数(即校验位)。
- ③ 选取一个 $K+1$ 位的产生多项式 $G(X)$,对 $M(X) \times X^K$ 作模 2 除。
- ④ 把左移 K 位以后的有效信息与余数 $R(X)$ 作模 2 加减,拼接为 CRC 码,此时的循环冗余校验码共有 $N+K$ 位。

例 2-4 已知 $M(X)=1001$, $G(X)=1011$,试计算出校验位,并组成循环冗余校验码。

解: 因为 $G(X)$ 有 4 位,所以 $K=3$ 。将 $M(X)$ 左移 3 位,得到 1001000。

进行模 2 除, $1001000 \div 1011 = 1010$,余数为 110,即循环冗余检验位。

$$\begin{array}{r}
 1010 \text{ ——商} \\
 1011 \overline{) 1001000} \\
 \underline{1011} \\
 0100 \\
 \underline{0000} \\
 1000 \\
 \underline{1011} \\
 0110 \\
 \underline{0000} \\
 110 \text{ ——余数}
 \end{array}$$

把余数加到 $M(X)$ 的后面,得到 1001**110**,此即循环冗余校验码。

在读出校验时,如果读出的校验码无误,那么 $1001110 \div 1011$,余数应为 0。

$$\begin{array}{r}
 1010 \text{ —— 商} \\
 1011 \overline{) 1001110} \\
 \underline{1011} \\
 0101 \\
 \underline{0000} \\
 1011 \\
 \underline{1011} \\
 0000 \\
 \underline{0000} \\
 000 \text{ —— 余数}
 \end{array}$$

当出错时,余数不为 0。假设 1001110 误作 1000110,余数为 011,将该余数的基础上添 0 后继续进行模 2 除法,余数的循环次序如图 2-15 所示。

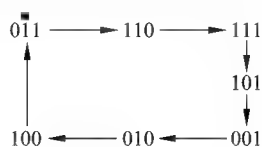


图 2-15 (7,4)码余数的循环次序

11. 循环冗余校验码的纠错原理

现用 A_7, A_6, A_5, A_4 表示代码的信息, Q_4, Q_3, Q_2, Q_1 表示 4 位商, A_3, A_2, A_1 表示余数,对于特定的生成多项式 $G(X)=1011$,列出算式如图 2-16 所示。

$$\begin{array}{r}
 \begin{array}{cccccccc}
 & & & & Q_4 & Q_3 & Q_2 & Q_1 & \text{—— 商} \\
 1011 \overline{) \begin{array}{cccccccc} A_7 & A_6 & A_5 & A_4 & 0 & 0 & 0 & 0 \end{array}} \\
 \hline
 \begin{array}{cccccccc} Q_4 & 0 & Q_4 & Q_4 & & & & \end{array} \\
 \hline
 \begin{array}{cccccccc} x & \Phi & \Phi & \Phi & 0 & & & \end{array} \\
 \hline
 \begin{array}{cccccccc} Q_3 & 0 & Q_3 & Q_3 & & & & \end{array} \\
 \hline
 \begin{array}{cccccccc} y & \Phi & \Phi & \Phi & 0 & & & \end{array} \\
 \hline
 \begin{array}{cccccccc} Q_2 & 0 & Q_2 & Q_2 & & & & \end{array} \\
 \hline
 \begin{array}{cccccccc} z & \Phi & \Phi & \Phi & 0 & & & \end{array} \\
 \hline
 \begin{array}{cccccccc} Q_1 & 0 & Q_1 & Q_1 & & & & \end{array} \\
 \hline
 \begin{array}{cccccccc} & & A_3 & A_2 & A_1 & & & \text{—— 余数} \end{array}
 \end{array}
 \end{array}$$

图 2-16 除法算式

因为部分余数的最高位为 1 则商为 1,最高位为 0 则商为 0,所以 $Q_4 = A_7$

上面的算式中,如 x 为 1,则 $Q_3=1$; x 为 0,则 $Q_3=0$,所以

$$Q_3=x=A_6\oplus 0=A_6$$

同理

$$Q_2=y=A_5\oplus Q_4=A_5\oplus A_7$$

$$Q_1=z=A_4\oplus Q_5\oplus Q_3=A_4\oplus A_7\oplus A_6$$

从算式中还可看出

$$A_3=Q_3\oplus Q_2=A_6\oplus A_5\oplus A_7$$

$$A_2=Q_2\oplus Q_1=A_4\oplus A_5\oplus A_6$$

$$A_1=Q_1=A_4\oplus A_7\oplus A_6$$

从以上 3 个表达式可以看出：

- A_3 、 A_5 、 A_6 、 A_7 组成一偶校验组,而 A_3 是它的校验位；
- A_2 、 A_4 、 A_5 、 A_6 组成一偶校验组,而 A_2 是它的校验位；
- A_1 、 A_4 、 A_6 、 A_7 组成一偶校验组,而 A_1 是它的校验位。

表 2-6 列出循环码的校验组。

表 2 6 循环码的校验组

	A_7	A_6	A_5	A_4	A_3	A_2	A_1
组 1	✓	✓	✓		✓		
组 2		✓	✓	✓		✓	
组 3	✓	✓		✓			✓

由此可以看出,循环冗余校验码编码原理与海明码相同,只是方法不同。循环冗余校验码在存储和传送过程中若是出现错误,则它除以原生成多项式后余数不等于零。根据表 2-5,很容易找到出错位和余数的对应关系,如表 2-7 所示。

表 2 7 出错位和余数的对应关系

出错位	影响	余数(出错模式)
A_7	A_3 、 A_1	101
A_6	A_3 、 A_2 、 A_1	111
A_5	A_3 、 A_2	110
A_4	A_2 、 A_1	011
A_3	A_3	100
A_2	A_2	010
A_1	A_1	001

它的出错顺序恰好是出错模式一栏从下往上的顺序,达到顶端以后又折回底部循环。如果把出错码从左往右计算位数,则左起第 1 位(A_7),出错时余数为 101;第 2 位(A_6)出错时余数为 111,把 111 继续用 1011 去除,则除一次后即得到 101;第 3 位(A_5)出错,则把余数连除两次,即得到 101。如左起第 n 位出错,则余数除 $n-1$ 次后得到 101。从这一规律可得出一个简单的纠错方法。即

- ① 余数为 0 时表示无错。
- ② 余数为 101 时,左起第一位出错。
- ③ 余数非 0 又非 101 时,继续模 2 除法。设除 $n-1$ 次得到余数 101,则从左起 n 位出错。

2.4 教材习题解答

1. 设机器数的字长 8 位(含 1 位符号位),分别写出下列各二进制数的原码、补码和反码:0, -0, 0.1000, -0.1000, 0.1111, -0.1111, 1101, -1101。

解:

真 值	原 码	补 码	反 码
0	00000000	00000000	00000000
-0	10000000	00000000	11111111
0.1000	0.1000000	0.1000000	0.1000000
-0.1000	1.1000000	1.1000000	1.0111111
0.1111	0.1111000	0.1111000	0.1111000
-0.1111	1.1111000	1.0001000	1.0000111
1101	00001101	00001101	00001101
-1101	10001101	11110011	11110010

2. 写出下列各数的原码、补码和反码: $\frac{7}{16}, \frac{4}{16}, \frac{1}{16}, \pm 0, -\frac{1}{16}, -\frac{4}{16}, -\frac{7}{16}$ 。

解: $\frac{7}{16} = 7 \times 2^{-4} = 0.0111$

$\frac{4}{16} = 4 \times 2^{-4} = 0.0100$

$\frac{1}{16} = 1 \times 2^{-4} = 0.0001$

真 值	原 码	补 码	反 码
$\frac{7}{16}$	0.0111	0.0111	0.0111
$\frac{4}{16}$	0.0100	0.0100	0.0100
$\frac{1}{16}$	0.0001	0.0001	0.0001
0	0.0000	0.0000	0.0000
-0	1.0000	0.0000	1.1111
$-\frac{1}{16}$	1.0001	1.1111	1.1110
$-\frac{4}{16}$	1.0100	1.1100	1.1011
$-\frac{7}{16}$	1.0111	1.1001	1.1000

3. 已知下列数的原码表示,分别写出它们的补码表示: $[X_1]_{\text{原}}=0.10100$, $[X_2]_{\text{原}}=1.10111$ 。

解: $[X_1]_{\text{补}}=0.10100$, $[X_2]_{\text{补}}=1.01001$ 。

4. 已知下列数的补码表示,分别写出它们的真值: $[X_1]_{\text{补}}=0.10100$, $[X_2]_{\text{补}}=1.10111$ 。

解: $X_1=0.10100$, $X_2=-0.01001$ 。

5. 设一个二进制小数 $X \geq 0$,表示成 $X=0.a_1a_2a_3a_4a_5$,其中 $a_i \sim a_5$ 取“1”或“0”:

(1) 若要 $X > \frac{1}{2}$, $a_1 \sim a_5$ 要满足什么条件?

(2) 若要 $X \geq \frac{1}{8}$, $a_1 \sim a_5$ 要满足什么条件?

(3) 若要 $\frac{1}{4} \geq X > \frac{1}{16}$, $a_1 \sim a_5$ 要满足什么条件?

解: (1) $X > \frac{1}{2}$ 的代码为:0.100001~0.111111。

$a_1=1, a_2+a_3+a_4+a_5=1$ 。

(2) $X \geq \frac{1}{8}$ 的代码为:

0.001000

...

0.111111

$\frac{1}{8}$

 $\frac{63}{64}$

$a_1+a_2=0, a_3=1$ 或 $a_1=0, a_2=1$ 或 $a_1=1$ 。

(3) $\frac{1}{4} \geq X > \frac{1}{16}$ 的代码为:

0.000101 $\frac{5}{64}$

...

0.010000 $\frac{1}{4}$

$a + a + a = 0, a = 1, a + a = 1$ 或 $a + a = 0, a = 1$ 或 $a = 1, a + a + a + a + a = 0$ 。

6. 设 $[X]_{\text{原}} = 1.a_1 a_2 a_3 a_4 a_5$,

(1) 若要 $X > -\frac{1}{2}$, $a \sim a_5$ 要满足什么条件?

(2) 若要 $-\frac{1}{8} \geq X \geq -\frac{1}{4}$, $a \sim a_5$ 要满足什么条件?

解: (1) $X > -\frac{1}{2}$ 的代码为:

1.000001 $-\frac{1}{64}$

...

1.011111 $-\frac{31}{64}$

$a = 0, a + a + a + a + a = 1$ 。

(2) $-\frac{1}{8} \geq X \geq -\frac{1}{4}$ 的代码为:

1.001000 $-\frac{1}{8}$

1.001001 $-\frac{9}{64}$

...

1.001111 $-\frac{15}{64}$

1.010000 $-\frac{1}{4}$

$a + a = 0, a = 1$ 或 $a = 1, a + a + a + a + a = 0$ 。

7. 若上题中 $[X]_{\text{原}}$ 改为 $[X]_{\text{补}}$, 结果如何?

解: 设 $[X]_{\text{补}} = 1.a_1 a_2 a_3 a_4 a_5$,

(1) $X > -\frac{1}{2}$ 的代码为:

$$\begin{array}{rcl} 1.100001 & - & \frac{31}{64} \\ \dots & & \\ 1.111111 & - & \frac{1}{64} \end{array}$$

$a=1, a+a+a+a+a=1。$

(2) $-\frac{1}{8} \geq X \geq -\frac{1}{4}$ 的代码为：

$$\begin{array}{rcl} 1.110000 & - & \frac{1}{4} \\ 1.110001 & - & \frac{15}{64} \\ \dots & & \\ 1.110111 & - & \frac{9}{64} \\ 1.111000 & - & \frac{1}{8} \end{array}$$

$a \cdot a=1, a=0$ 或 $a \cdot a \cdot a=1, a+a+a=0。$

8. 一个 n 位字长的二进制定点整数,其中 1 位为符号位,分别写出在补码和反码两种情况下:

- (1) 模数;
- (2) 最大的正数;
- (3) 最负的数;
- (4) 符号位的权;
- (5) -1 的表示形式;
- (6) 0 的表示形式。

解:

项目	补码	反码
模数	$\text{Mod } 2^n$	$\text{Mod } (2^n-1)$
最大的正数	$2^{n-1}-1$	$2^{n-1}-1$
最负的数	-2^{n-1}	$-(2^{n-1}-1)$
符号位的权	2^{n-1}	2^{n-1}
-1 的表示形式	11111111	11111110
0 的表示形式	00000000	00000000 11111111

9. 某机字长 16 位,问在下列几种情况下所能表示数值的范围:

- (1) 无符号整数;
- (2) 用原码表示定点小数;
- (3) 用补码表示定点小数;
- (4) 用原码表示定点整数;
- (5) 用补码表示定点整数。

解: (1) $0 \leq X \leq (2^{16} - 1)$
 (2) $-(1 - 2^{-15}) \leq X \leq (1 - 2^{-15})$
 (3) $-1 \leq X \leq (1 - 2^{-15})$
 (4) $-(2^{15} - 1) \leq X \leq (2^{15} - 1)$
 (5) $-2^{15} \leq X \leq (2^{15} - 1)$

10. 某机字长 32 位, 试分别写出无符号整数和带符号整数(补码)的表示范围(用十进制数表示)。

解: 无符号整数: $0 \leq X \leq (2^{32} - 1)$ 。

补码: $-2^{31} \leq X \leq (2^{31} - 1)$ 。

11. 某浮点数字长 12 位, 其中阶符 1 位, 阶码数值 3 位, 数符 1 位, 尾数数值 7 位, 阶码以 2 为底, 阶码和尾数均用补码表示。它所能表示的最大正数是多少? 最小规格化正数是多少? 绝对值最大的负数是多少?

解: 最大正数 $= (1 - 2^{-7}) \times 2^{2^3 - 1} = (1 - 2^{-7}) \times 2^7 = 127$ 。

最小规格化正数 $= 2^{-1} \times 2^{-2^3} = 2^{-1} \times 2^{-8} = 2^{-9} = \frac{1}{512}$ 。

绝对值最大的负数 $= -1 \times 2^{2^3 - 1} = -1 \times 2^7 = -128$ 。

12. 某浮点数字长 16 位, 其中阶码部分 6 位(含 1 位阶符), 移码表示, 以 2 为底; 尾数部分 10 位(含 1 位数符, 位于尾数最高位), 补码表示, 规格化。分别写出下列各题的二进制代码与十进制真值。

- (1) 非零最小正数;
- (2) 最大正数;
- (3) 绝对值最小负数;
- (4) 绝对值最大负数。

解: (1) 非零最小正数: 000000, 0, 100000000; $2^{-1} \times 2^{-2^5} = 2^{-33}$ 。

(2) 最大正数: 111111, 0, 111111111; $(1 - 2^{-9}) \times 2^{2^5 - 1} = (1 - 2^{-9}) \times 2^{31}$ 。

(3) 绝对值最小负数: 000000, 1, 011111111; $-(2^{-1} + 2^{-9}) \times 2^{-2^5}$ 。

(4) 绝对值最大负数: 111111, 1, 000000000; $-1 \times 2^{2^5 - 1} = -2^{31}$ 。

13. 一浮点数, 其阶码部分为 p 位, 尾数部分为 q 位, 各包含 1 位符号位, 均用补码表示; 尾数基数 $r=2$, 该浮点数格式所能表示数的上限、下限及非零的最小正数是多少? 写

出表达式。

解：上限(最大正数) = $(1 - 2^{-(q-1)}) \times 2^{2^{(p-1)}-1}$ 。

下限(绝对值最大负数) = $-1 \times 2^{2^{(p-1)}-1}$ 。

最小正数 = $2^{-(q-1)} \times 2^{-2^{(p-1)}}$ 。

最小规格化正数 = $2^{-1} \times 2^{-2^{(p-1)}}$ 。

14. 若上题尾数基数 $r=16$, 按上述要求写出表达式。

解：上限(最大正数) = $(1 - 2^{-(q-1)}) \times 16^{2^{(p-1)}-1}$ 。

下限(绝对值最大负数) = $-1 \times 16^{2^{(p-1)}-1}$ 。

最小正数 = $2^{-(q-1)} \times 16^{-2^{(p-1)}}$ 。

最小规格化正数 = $16^{-1} \times 16^{-2^{(p-1)}}$ 。

15. 某浮点数字长 32 位, 格式如下。其中阶码部分 8 位, 以 2 为底, 移码表示; 尾数部分一共 24 位(含 1 位数符), 补码表示。现有一浮点代码为 $(8C5A3E00)_{16}$, 试写出它所表示的十进制真值。

0	7	8	9	3
阶码		数符	尾数	

解： $(8C5A3E00)_{16} = 1000\ 1100\ 0101\ 1010\ 0011\ 1110\ 0000\ 0000B$,

$0.10110100011111 \times 2^{12} = (101101000111.11)_2 = (2887.75)_{10}$ 。

16. 试将 $(-0.1101)_2$ 用 IEEE 短浮点数格式表示出来。

解： $0.1101 = 1.101 \times 2^{-1}$ 。

符号位 = 1。

阶码 = $127 - 1 = 126$ 。

1, 01111110, 101000000000000000000000。

结果 = BF500000H。

17. 将下列十进制数转换为 IEEE 短浮点数：

(1) 28.75;

(2) 624;

(3) -0.625 ;

(4) $+0.0$;

(5) -1000.5 。

解：(1) $28.75 = 11100.11 = 1.110011 \times 2^4$ 。

符号位 = 0。

阶码 = $127 + 4 = 131$ 。

0, 10000011, 110011000000000000000000。

结果=41E60000H。

(2) $624 = 1001110000 = 1.001110000 \times 2^9$ 。

符号位=0。

阶码=127+9=136。

0,10001000,001110000000000000000000。

结果=441C0000H。

(3) $-0.625 = -0.101 = -1.01 \times 2^{-1}$ 。

符号位=1。

阶码=127-1=126。

1,01111110,010000000000000000000000。

结果=BF200000H。

(4) +0.0。

结果=00000000H。

(5) $-1000.5 = 111101000.1 = 1.111010001 \times 2^9$ 。

符号位=1。

阶码=127+9=136。

1,10001000,111010001000000000000000。

结果=C47A2000H。

18. 将下列 IEEE 短浮点数转换为十进制数：

(1) 11000000 11110000 00000000 00000000；

(2) 00111111 00010000 00000000 00000000；

(3) 01000011 10011001 00000000 00000000；

(4) 01000000 00000000 00000000 00000000；

(5) 01000001 00100000 00000000 00000000；

(6) 00000000 00000000 00000000 00000000。

解：(1) 1,10000001,111000000000000000000000

符号位=1。

阶码=129-127=2。

$1.111 \times 2^2 = 111.1B = 7.5$ 。

所以结果=-7.5。

(2) 0,01111110,001000000000000000000000

符号位=0。

阶码=126-127=-1。

$1.001 \times 2^{-1} = 0.1001B = 0.5625$ 。

所以,结果=0.5625。

(3) 0,10000111,001100100000000000000000

符号位=0。

阶码=135-127=8。

$1.0011001 \times 2^8 = 100110010\text{B} = 306$ 。

所以,结果=306。

(4) 0,10000000,000000000000000000000000

符号位=0。

阶码=128-127=1。

$1.0 \times 2^1 = 10\text{B} = 2$ 。

所以,结果=2。

(5) 0,10000010,0100000 00000000 00000000

符号位=0。

阶码=130-127=3。

$1.01 \times 2^3 = 1010\text{B} = 10$ 。

所以,结果=10。

(6) 0,00000000,000000000000000000000000

阶码和尾数都等于全0,结果=0。

19. 对下列 ASCII 码进行译码:

1001001,0100001,1100001,1110111

1000101,1010000,1010111,0100100

解: 以上 ASCII 码分别为 I, !, a, w, E, P, W, \$。

20. 以下列形式表示 $(5382)_{10}$ 。

(1) 8421 码; (2) 余 3 码;

(3) 2421 码; (4) 二进制数。

解: (1) 0101 0011 1000 0010。

(2) 1000 0110 1011 0101。

(3) 1011 0011 1110 0010。

(4) 1010100000110。

21. 填写下列代码的奇偶校验位,现设为奇校验:

1 0 1 0 0 0 0 1

0 0 0 1 1 0 0 1

0 1 0 0 1 1 1 0

解: 3 个代码的校验位分别是 0, 0, 1。

22. 已知下面数据块约定:横向校验、纵向校验均为奇校验,请指出至少有多少位出错。

	a	b	c	d	e	f	g	h	校验位
	1	0	0	1	1	0	1	1	→ 0
	0	0	1	1	0	1	0	1	→ 1
	1	1	0	1	0	0	0	0	→ 0
	1	1	1	0	0	0	0	0	→ 0
	0	1	0	0	1	1	1	1	→ 0
	↓	↓	↓	↓	↓	↓	↓	↓	
校验位	1	0	1	0	1	1	1	1	

解:经检测 a 和 h 列出错,所以至少有两位出错。

23. 求有效信息位为 01101110 的海明校验码。

解: P_5 D8 D7 D6 D5 P_4 D4 D3 D2 P_3 D1 P_2 P_1

$$P_1 = D_1 \oplus D_2 \oplus D_4 \oplus D_5 \oplus D_7 = 0 \oplus 1 \oplus 1 \oplus 0 \oplus 1 = 1$$

$$P_2 = D_1 \oplus D_3 \oplus D_4 \oplus D_6 \oplus D_7 = 0 \oplus 1 \oplus 1 \oplus 1 \oplus 1 = 0$$

$$P_3 = D_2 \oplus D_3 \oplus D_4 \oplus D_8 = 1 \oplus 1 \oplus 1 \oplus 0 = 1$$

$$P_4 = D_5 \oplus D_6 \oplus D_7 \oplus D_8 = 0 \oplus 1 \oplus 1 \oplus 0 = 0$$

$$P_5 = D_1 \oplus D_2 \oplus D_3 \oplus D_5 \oplus D_6 \oplus D_8 = 0 \oplus 1 \oplus 1 \oplus 0 \oplus 1 \oplus 0 = 1$$

所以,海明校验码=1011001111001。

24. 设计算机准备传送的信息是:1010110010001111,生成多项式是 $X^5 + X^2 + 1$,计算校验位,写出 CRC 码。

解:生成多项式 $X^5 + X^2 + 1 = 100101$ 。

首先将准备传送的信息左移 5 位:101011001000111100000。

然后 $101011001000111100000 \div 100101$,余数=10011。

所以,CRC 码=101011001000111110011。

第 3 章

指令系统

3.1 基本内容要求

指令、指令系统是计算机中一个最基本的概念。指令是指示计算机执行某些操作的命令,一台计算机的所有指令的集合构成该机的指令系统。指令系统是计算机的主要属性,位于硬件和软件的交界面上。本章将讨论一般计算机的指令系统所涉及的基本问题。

学习要求

- ◆ 了解指令的基本格式
- ◆ 理解不同地址码(三、二、一、零地址)双操作数指令的区别
- ◆ 理解规整型指令(定长操作码)的特点
- ◆ 理解非规整型指令(扩展操作码)的特点
- ◆ 掌握扩展操作码指令的格式设计
- ◆ 了解编址单位的概念及常见的编址(字编址和字节编址)计算机的特点
- ◆ 理解指令中地址码的位数与主存容量、最小寻址单位的关系
- ◆ 理解数据寻址和指令寻址的区别
- ◆ 了解数据寻址的最终目的
- ◆ 理解常见寻址方式(立即寻址、直接寻址、寄存器寻址、间接寻址、寄存器间接寻址、变址寻址、相对寻址、页面寻址)的特点
- ◆ 掌握直接寻址、间接寻址、变址寻址、相对寻址和页面寻址方式中有效地址 EA 的计算
- ◆ 理解自底向上的存储器堆栈的概念及堆栈的进、出栈操作
- ◆ 掌握进栈、出栈时栈指针的修改和数据的压入和弹出
- ◆ 理解转移、转子、返回指令的特点与区别
- ◆ 理解独立编址 I/O 和统一编址 I/O 的区别

3.2 误点疑点解惑

1. 指令长度

指令长度即一条指令中包含的二进制代码的位数,是指令格式设计最基本的出发点。为了便于处理字符数据,尽可能充分利用存储空间,减少指令和数据的存储时间,现代计算机的机器字长、存储器宽度和 I/O 传输宽度几乎都是字节的整倍数,这个限制同样适合于指令长度。

从访问存储器和指令的复杂性的角度来看,短指令比长指令好。短指令能够节省存储空间,减少访问存储器次数,具有快的执行速度。指令越短,意味着占用的存储器规模(指令长度 \times 指令数)就越小,这是显而易见的。而减少访存次数、提高指令的执行速度是体现在增加了单位时间内取出指令的条数上,若存储器传送速率为 T 位/秒,指令平均长度为 L 位,则每秒传送指令数为 T/L 条, L 越小,则 T/L 就越大,单位时间内从存储器中取出的指令条数就越多。在目前的存储技术条件下,取指令比执行指令时间要大得多,故短指令有利于提高计算机的执行速度。但短指令也有其不可克服的局限性,这就是指令中包括的信息少,指令功能较弱。现代计算机仍然希望发展长指令,长指令的功能比较强,便于程序设计,但长指令又可能造成利用率不高的浪费。为了合理地安排存储空间,并使指令能表达较丰富的含义,通常指令系统采用变长指令字结构。例如,IBM 370 的指令长度有 16 位(半字)的,有 32 位(单字)的,有 48 位(一个半字)的。Pentium 也是变长指令字结构,指令包括从 8~128 位的多种形式。变长指令字结构使用灵活,能充分利用指令长度,但指令控制较复杂。当采用变长指令格式时,往往将操作码放在第一字节中,用以判明该指令的基本类型及相应字节数。而且,通常把最常用的指令设计成短指令,以便节省存储空间和提高指令的执行速度。对于变长指令字结构,指令长度并不是任意的。为了充分利用存储空间,指令长度通常为字节的整倍数,以避免存储空间的浪费。

目前还有一种适合超大规模集成电路实现的计算机指令结构,称为超长指令字(very long instruction word, VLIW)。这种结构的指令字长度在 100 位以上,在一个指令字中包含一组多种类型并可以同时执行的指令,借助于集成电路技术的支持,在 CPU 中设计大量的功能部件同时执行这一组指令,使系统达到很高的性能。

2. 双操作数运算类指令的执行

对于双操作数运算类指令(如加法指令)来说,每条指令中都需要包含以下 4 个地址信息:

- ① 第一操作数地址 A_1 ;
- ② 第二操作数地址 A_2 ;
- ③ 操作结果存放地址 A_3 ;

④ 下条将要执行指令的地址 A_4 。

这些地址信息可以明显地给出,称为显地址;也可以依照某种事先的约定,用隐含的方式给出,称为隐地址。

大多数计算机中用程序计数器 PC 指出下一条将要执行指令的地址。PC 是一个兼有寄存器和计数器功能的部件,它指引着 CPU 从何处去读取指令。通常,程序往往存放在主存的一段连续区域中,CPU 在执行完一条指令后,PC 的内容计数($PC+1$),指出下条指令的地址,所以指令中不需要明显地给出 A_4 。除去隐含约定的地址 A_4 以外,其余 3 个地址的处理方式有 4 种。

(1) 三地址双操作数指令

三地址双操作数指令有 3 个显地址,指令的含义为

$$(A_1)OP(A_2) \rightarrow A_3$$

假设指令存放在主存的 50 号单元中,第一、第二操作数分别存放在主存的 100 和 200 号单元中,结果存放在主存的 300 号单元,如图 3-1 所示。

执行一条三地址的加法指令需要访问 4 次主存。第一次从 50 号单元中取指令,第二次从 100 号单元中取第一操作数,第三次从 200 号单元中取第二操作数,第四次将加法的结果保存到主存的 300 号单元。

(2) 二地址双操作数指令

二地址双操作数指令有两个显地址,第一操作数地址同时兼作结果存放地址(目的地址),指令的含义为

$$(A_1)OP(A_2) \rightarrow A_1$$

执行一条二地址的加法指令同样需要访问 4 次主存。第一次从 50 号单元中取指令,第二次从 100 号单元中取第一操作数,第三次从 200 号单元中取第二操作数,第四次将加法的结果保存到主存的 100 号单元。

(3) 一地址双操作数指令

一地址双操作数指令只有一个显地址,参加运算的另一个操作数来自累加寄存器 Acc。指令的含义为

$$(Acc)OP(A_1) \rightarrow Acc$$

执行一条一地址的加法指令只需要访问两次主存,第一次从 50 号单元中取指令,第二次从 100 号单元中取操作数。由于第一操作数和运算结果都放在累加寄存器中,所以读取第一操作数和存放加法的结果都不需要访问主存。

(4) 零地址双操作数指令

零地址双操作数指令中只有操作码字段,操作数地址都是隐含的。操作数在堆栈的

50	指令
51	下一指令
	⋮
100	第一操作数
	⋮
200	第二操作数
	⋮
300	
	⋮

图 3-1 存放在主存中的指令和数据

栈顶位置和次栈顶位置,它们分别从堆栈中弹出,送到运算器中进行运算,运算的结果再压入堆栈。

执行一条零地址的加法指令访问主存的次数取决于堆栈的结构。如果是软堆栈,则需要访问主存 4 次,因为软堆栈就是主存的一部分;如果是硬堆栈,则只需要访存 1 次,因为硬堆栈是由寄存器组成的。

最后要特别指出的是,前面提到的 $PC+1$ 中的“1”实际上是指一个增量,并不一定是数值 1。对于一个字节编址的计算机来说,假定一条指令只占一个字节,则 PC 内容+1 指向下一条指令地址,假定一条指令占 n 个字节,则 PC 的内容+ n 指向下一条指令地址。

3. 地址个数对程序长度和指令长度的影响

从缩短程序长度、用户使用方便、增加操作并行度等方面来看,选用三地址指令较好;从缩短指令长度、减少访存次数、简化硬件设计等方面来看,一地址指令较好。对于同一个问题,用三地址指令编写的程序最短,但指令长度最长,而用二、一、零地址指令来编写程序,程序的长度一个比一个长,但指令的长度一个比一个短。

例 3-1 分别利用三地址、二地址、一地址和零地址指令编制计算算术表达式 $x=(a \times b+c-d) \div (e+f)$ 的程序。

解: 假设 a, b, c, \dots 为操作数, A, B, C, \dots 为操作数地址。

(1) 三地址指令程序

```
MUL    A,B,X
ADD     X,C,X
SUB     X,D,X
ADD     E,F,Y
DIV     X,Y,X
```

共需 5 条三地址指令,每条指令 4 次访存,执行此程序共访存 20 次。

(2) 二地址指令程序

```
MOV     X,A
MUL     X,B
ADD     X,C
SUB     X,D
MOV     Y,E
ADD     Y,F
DIV     X,Y
```

共需 7 条二地址指令,MOV 指令 3 次访存,算术运算指令 4 次访存,执行此程序共访存 $2 \times 3 + 5 \times 4 = 26$ 次。

(3) 一地址指令程序

```
LOAD    E
ADD     F
STORE   X
LOAD    A
MUL     B
ADD     C
SUB     D
DIV     X
STORE   X
```

共需 9 条一地址指令,每条指令 2 次访存,执行此程序共访存 $9 \times 2 = 18$ 次。

(4) 零地址指令程序

```
PUSH    A
PUSH    B
MUL
PUSH    C
ADD
PUSH    D
SUB
PUSH    E
PUSH    F
ADD
DIV
POP     X
```

共需 12 条指令,其中 7 条一地址的进、出栈指令,5 条零地址的算术运算指令。进、出栈指令 3 次访存,算逻指令 4 次访存,执行此程序共访存 $7 \times 3 + 5 \times 4 = 41$ 次。

大家可能已经注意到,在例 3-1 中故意回避了一个问题,即没有考虑指令本身的长度对取指令的影响,简单地认为所有的指令都是一次从主存中取出来的。事实上,由于不同地址数的指令的长度不同,如果要考虑它们从主存中取出的情况,问题会复杂不少。下面看另一个例子。

例 3-2 某一机器的指令系统,操作码 8 位,地址码均为 16 位,CPU 与主存之间每次传送 16 位数据。A、B、C、D、E 表示字地址,存放 16 位数据。

(1) 分别用二地址、二地址、一地址和零地址指令编写程序,计算 $A = (B - C) \times (D - E)$ (不允许覆盖任何操作数,可以使用暂存单元)。

(2) 分别计算所写程序的总字节数。

(3) 分别计算程序执行时的访存次数。

解：(1) 设暂存为 Tmp。则 4 段程序分别如下所示。

① 三地址指令程序

```
SUB    B,C,A
SUB    D,E,Tmp
MPY    A,Tmp,A
```

② 二地址指令程序

```
MOV    A,B
SUB    A,C
MOV    Tmp,D
SUB    Tmp,E
MPY    A,Tmp
```

③ 一地址指令程序

```
LOAD   D
SUB     E
STORE  Tmp
LOAD   B
SUB     C
MPY    Tmp
STORE  A
```

④ 零地址指令程序

```
PUSH   B
PUSH   C
SUB
PUSH   D
PUSH   E
SUB
MPY
POP    A
```

(2) 因为操作码 8 位,地址码 16 位,所以:

- 三地址指令程序中每条指令占 7 个字节,程序的字节总数为 21 个字节;
- 二地址指令程序中每条指令占 5 个字节,程序的字节总数为 25 个字节;
- 一地址指令程序中每条指令占 3 个字节,程序的字节总数为 21 个字节;
- 零地址指令程序需要有 5 条一地址指令(3 个字节),3 条零地址指令(1 个字节),

程序的字节总数为 18 个字节。

(3) 因为三地址指令占 7 个字节,每条指令需访存 4 次才能取出,接下来每条指令还需要访存 3 次,所以 3 条指令共访存 21 次。

因为二地址指令占 5 个字节,每条指令需访存 3 次才能取出。除取指令外,传送指令还需访存 2 次,运算指令还需访存 3 次,所以总的访存次数为: $5 \times 3 + 2 \times 2 + 3 \times 3 = 28$ 次。

因为一地址指令占 3 个字节,每条指令需访存 2 次才能取出。除取指令外,还需访存 1 次,所以总的访存次数为: $7 \times 2 + 7 = 21$ 次。

零地址指令占 1 个字节,每条指令访存 1 次即可取出。零地址指令程序中取指令访存次数为: $5 \times 2 + 3 \times 1 = 13$ 次,除取指令外,进、出栈指令还需访存 2 次,运算指令还需访存 3 次,所以总的访存次数为: $13 + 5 \times 2 + 3 \times 3 = 32$ 次。

4. 不同地址数指令的进一步分析

前面分析了双操作数运算类指令的不同地址数,需要提醒学生注意的是,这些指令都至少需要 3 个地址(不含下一指令地址),但是实际上在不同计算机的指令系统中,双操作数运算类指令的地址数可以不同,即有些机器中这 3 个地址可能均是显地址,也有些机器中可能只有 2 个显地址或 1 个显地址或根本就没有显地址(如堆栈计算机),所以就出现了双操作数运算类的三、二、一、零地址指令。

下面提到的不同地址数指令则是从另一个角度来讨论问题,即根据指令中实际需要的地址个数来决定指令的地址数。如双操作数的运算类指令需要 3 个地址,所以是三地址指令;而传送类指令需要两个地址,所以是二地址指令;单操作数的运算类指令(如+1、-1、求反等)当然只需要一个地址,所以是一地址指令;还有一些指令不需要操作数(如停机、空操作、清除等控制类指令),它们就是零地址指令。

前面讨论了执行一条指令访问主存的次数,应当提请学生注意,必须依据具体的指令所完成的操作来决定访问主存的次数,不要认为同一地址数的指令访问主存的次数都是一样的。

例如:传送类指令 MOV A,B 是一条二地址指令,执行这样的一条二地址指令和前述的二地址双操作数指令访问主存的次数是不同的。MOV 指令访问主存的次数只有 3 次,第一次取指令,第二次取源操作数,第三次将结果存放在目标地址内。

再比如:加 1 指令 INC A 是一条一地址指令,执行这样的一地址指令访问主存的次数也要 3 次。一地址单操作数运算类指令仅需要一个操作数,指令的含义为

$$OP(A_1) \rightarrow A_1$$

5. 非规整性编码 扩展操作码法

在一个计算机的指令系统中,不同的指令需要的地址个数是不相等的。假设指令系统中有下列几条指令:

```

ADD A,B,C
MOV A,B
INC A
HALT

```

以上 4 条指令分别为三、二、一、零地址,指令的地址码的长度随着地址码个数的增加而增加。

指令操作码的编码可以分为规整型和非规整型两类,最常用的非规整型编码方式是扩展操作码法。假设指令长度一定,则地址码与操作码字段的长度是相互制约的。让那些操作数地址个数多的指令(三地址指令)的操作码字段短些,让那些操作数地址个数少的指令(一或零地址指令)的操作码字段长些,这样既能充分地利用指令的各个字段,又能在不增加指令长度的情况下扩展操作码的位数,使它能表示更多的指令。

不论采用何种方案,必须要注意以下两点:

- 不允许短码是长码的前缀,即短操作码不能与长操作码的前面部分的代码相同,否则将无法保证解码的惟一性和实时性。
- 各条指令的操作码一定不能重复,而且各类指令的格式安排应统一规整。

也就是说,不能用已经定义过的操作码再作扩展窗口来扩展其他的指令。

6. 地址码位数与主存容量和最小寻址单位的关系

指令格式中每个地址码的位数是与主存容量和最小寻址单位有关联的。最小寻址单位就是编址单位,常见的编址单位有字编址和字节编址。主存容量越大,访问全部存储空间所需的地址码位数就越长,这是很容易理解的。当存储容量确定之后,如果主存采用字节编址,所需的地址码的位数就需要长些;如果主存采用字编址(假定字长为 16 位或更长),所需的地址码的位数就需要短些。这就如同一个大楼的总面积一定,若每个房间的面积小,则房间数就多,对应的地址位数就长;若每个房间的面积大,则房间数就少,对应的地址位数就短。

例 3-3 设某机主存容量为 16MB,机器字长 16 位,若最小寻址单位为字节(按字节编址),其地址码为多少位? 若最小寻址单位为字(按字编址),其地址码又为多少位?

解: 若按字节编址,地址码应为 24 位($2^{24} = 16\text{MB}$);若按字编址,地址码只需 23 位。这是因为 16 位的一个字等于两个字节,有

$$16\text{MB} = \frac{16\text{M}}{2} \text{W} = \frac{2^{24}}{2^1} \text{W} = 2^{23} \text{W} = 8\text{M W}$$

例 3-4 设某机为 32 位的 16MB 主存,若按字编址,其地址码为多少位? 若按字节编址,其地址码又为多少位?

解: 若按字编址,地址码应为 24 位,字长 32 位;若按字节编址,地址码应为 26 位。这是因为 32 位的一个字等于 4 个字节,有

$$16\text{MW}=4\times 16\text{MB}=2^2\times 2^{24}\text{B}=2^{26}\text{B}=64\text{MB}$$

在讲解这类例题时,首先要讲清楚主存容量与地址码的关系,这对初学者来说一般都会有些困难;然后要仔细分析字与字节的关系,即一个字由几个字节组成;最后就可以比较容易地推出结果了。

7. 指令寻址和数据寻址

寻址可以分为指令寻址和数据寻址。寻找下一条将要执行的指令地址称为指令寻址,寻找操作数的地址称为数据寻址。

在主教材中用比较大的篇幅来讨论数据寻址,但这并不代表指令寻址就不重要。指令寻址相对数据寻址来说,寻址方式比较简单,可以细分为顺序寻址和跳跃寻址。顺序寻址可通过程序计数器 PC 加增量,自动形成下一条指令的地址;跳跃寻址则需要通过程序转移类指令实现,其转移地址的形成方式有 3 种:直接(绝对)、相对和间接寻址。

8. 常见数据寻址方式分析

CPU 根据指令约定的寻址方式对地址字段的有关信息作出解释,以找到操作数。有的指令设置专门的寻址方式字段,以说明采用何种寻址方式,有的指令则通过操作码的含义,隐含约定采用何种寻址方式。对于涉及多个地址的指令,各个地址可以有各自的寻址方式,也就是说,一条指令中可以有多种寻址方式。

一个指令系统具有哪几种寻址方式,这是设计指令系统的关键,也是初学者理解一个指令系统的难点所在。因此,在这部分内容的教学过程中,应帮助学生从众多的寻址方式中归纳出一条清晰的思路来。

首先看指令要调用的操作数可能存放在什么地方? 经过分析发现,操作数所在的位置无非有下列 5 种情况:

- ① 操作数就包含在某指令中或紧跟着某指令,相应地需要由指令直接给出操作数。
- ② 操作数在 CPU 的某个寄存器中,相应地需要指令中给出寄存器编号。
- ③ 操作数在主存中,则指令应以某种方式给出主存单元的地址码。这里还可分为几种情况:有的是对单个操作数进行处理;有的是对一个连续的数组或对数组中的某个元素进行处理;有的是对一个表格或对表格中的某个元素进行处理等。需要相应地采取不同的寻址方式。
- ④ 操作数在堆栈区中,可以隐含约定由堆栈指针 SP 提供地址。
- ⑤ 操作数在某个 I/O 接口的寄存器中,指令中需要提供 I/O 端口地址(独立编址)或总线地址(统一编址)。

接下来,沿着从简到繁的思路,大致可将众多的寻址方式归纳为以下 4 大类:

- ① 立即寻址——在读取指令时从指令中获得操作数。
- ② 直接寻址类——直接给出寄存器编号或主存单元地址,以获取操作数(如寄存器寻址、直接寻址)。

③ 间接寻址类——先从某寄存器或主存中读取地址,再按这个地址访问主存,读取操作数(如寄存器间接寻址、间接寻址)。

④ 变址类 指令给出的是形式地址,经过某种计算(例如相加、相减、高低位地址拼接等),才获得有效地址,据此访问主存以读取操作数(如变址寻址、基址寻址、相对寻址、页面寻址)。

尽管各种计算机的寻址方式种类甚多,尤其是不同系列的计算机之间更是既有大体相同之处,也有各具特色之处。不同的计算机对寻址方式的分类和命名也有各自的规定,但几乎都是以上述4类为最基本的寻址方式,其他的则是它们的变型或组合。沿着上述思路去学习,可以更好地理解各种寻址方式的含义。

9. 各种数据寻址方式的速度比较

在主教材中,一共讨论了9种基本的数据寻址方式,数据寻址的最终目的是寻找所需要的操作数。操作数可以在主存中,也可以在寄存器中,甚至可以在堆栈中。各种不同的寻址方式获取操作数的速度是不相同的。9种基本的数据寻址方式获取操作数的顺序依次是:

立即寻址

寄存器寻址 $EA = R_i$

直接寻址 $EA = A$

寄存器间接寻址 $EA = (R_i)$

页面寻址 $EA = PC // A$

变址寻址 $EA = (R_x) + A$

基址寻址 $EA = (R_b) + A$

相对寻址 $EA = (PC) + A$

间接寻址 $EA = (A)$

其中变址寻址、基址寻址和相对寻址又可以统称为偏移寻址,这几种寻址方式形成有效地址 EA 的机制相同,都是将指定寄存器的内容与指令中的地址码字段相加,所以获取操作数的速度相同。

直接寻址、寄存器间接寻址、页面寻址、变址寻址、基址寻址、相对寻址等获取操作数都只需要访问一次主存(不含取指令本身),根据有效地址 EA 得到的难易程度,速度上稍有差别。寄存器间接寻址由于要先到寄存器中取出操作数的地址,所以获取操作数的速度要稍慢于直接寻址,而页面寻址的有效地址 EA 通过简单的拼接得到,将稍快于变址寻址、基址寻址和相对寻址。

间接寻址指令中给出的地址 A 不是操作数的地址而是操作数地址的地址。这就意味着为获取一个操作数,至少需要两次访问主存(不含取指令本身)。

注:在主教材中,基址寻址和相对寻址的有效地址公式里使用的是字母 D 而不是字

母 A,其实这并不矛盾,因为无论是位移量(常用字母 D 表示)还是形式地址(常用字母 A)表示,都是由指令的地址码字段给出的,可以认为它们没有区别。

例 3-5 设寄存器 R 中的数值为 1000H,地址为 1000H 的存储单元中存储的内容为 2000H,地址为 2000H 的存储单元中存储的内容为 3000H,PC 的值为 4000H,问以下寻址方式下访问到的操作数的值是什么?

- (1) 寄存器寻址 R;
- (2) 寄存器间接寻址(R);
- (3) 直接寻址 1000H;
- (4) 间接寻址(1000H);
- (5) 相对寻址 -2000H(PC);
- (6) 立即寻址 #2000H。

解: (1) 采用寄存器寻址,操作数在寄存器中, $S = (R) = 1000H$ 。

(2) 采用寄存器间接寻址,操作数的有效地址在寄存器中, $EA = (R)$,操作数 $S = ((R)) = (1000H) = 2000H$ 。

(3) 采用直接寻址,操作数的有效地址在指令中给出, $EA = 1000H$,操作数 $S = (1000H) = 2000H$ 。

(4) 采用间接寻址,操作数有效地址在主存单元中, $EA = (1000H) = 2000H$,操作数 $S = (2000H) = 3000H$ 。

(5) 采用相对寻址,操作数有效地址为 PC 中的内容与指令中的位移量 D 之和,即 $EA = (PC) + D = 4000H - 2000H = 2000H$,操作数 $S = (2000H) = 3000H$ 。

(6) 采用立即寻址,操作数直接在指令中给出, $S = 2000H$ 。

10. 变址寻址和基址寻址的区别

变址寻址的含义是:指令中的地址部分给出一个形式地址,并且指定一个寄存器作为变址寄存器;变址寄存器的内容(变址值)与形式地址相加,得到操作数有效地址;按照有效地址访问主存,从相应的主存单元中读出操作数或向该单元写入数据。

变址方式的应用广泛,最典型的用法是将形式地址作为基准地址,例如某个数组的首址;变址寄存器的内容是修改值,又称变址值,它是访问单元与首址单元之间的距离。如果按照这种含义,则形式地址的位数应能提供全字长的地址码,可以覆盖整个存储空间;而变址寄存器提供的变址值,位数可以少些,只需覆盖操作对象所在区间即可。例如某计算机主存容量 64KB,作为操作对象的数据块(小于 256B)可以存放于主存的任一区间,按上述应用方式,形式地址应有 16 位,而变址值只需 8 位。当然上述应用方式也并非一成不变,可以根据实际需要灵活变化,例如在定长指令格式中,形式地址往往不能提供全字长的地址码,而变址寄存器的位数反而可能提供全字长的地址码。

基址寻址的含义是:指令中的地址部分给出一个形式地址,作为位移量,并且还指定

一个寄存器作为基址寄存器;基址寄存器的内容与形式地址相加,得到操作数有效地址;按照有效地址访问主存,从相应的主存单元中读出操作数或向该单元写入数据。

基址寻址的典型应用有两个:一是程序重定位,即由操作系统给用户程序分配一个基地址,并且将它装入基址寄存器,在执行程序时就可以自动形成实际的主存地址。二是扩展有限字长指令的寻址空间,即在运行时将某个主存区间的首址或程序段的首址装入基址寄存器,以便直接访问大容量主存的任一区间。例如,主存容量 16MB,基址寄存器 24 位,14 位形式地址给出位移量,可以访问某个 16KB 的空间,通过更改基址寄存器的内容,可以移向另一区间,各个存储区间可以部分重叠。

可见,虽然变址寻址和基址寻址形成有效地址的方法几乎相同,但具体应用不同。变址寻址立足于面向用户,可用于访问字符串、数组、表格等成批数据或其中的某些元素;基址寻址立足于面向系统,用来解决程序在实际主存中的重定位问题以及在有限字长指令中扩大寻址空间等。从使用方式看,在使用变址寻址时,一般由指令提供的形式地址作为基准地址,变址寄存器提供修改量;在使用基址寻址时,一般由基址寄存器提供基准地址,指令提供的形式地址作为位移量。当然,在实际机器中它们的具体应用方式可有不同的变化。

11. 相对寻址中的位移量

相对寻址由程序计数器 PC 提供基准地址,指令中的地址码字段作为位移量 D,两者相加后得到操作数的有效地址,即 $EA = (PC) + D$ 。位移量指出的是操作数和现行指令之间的相对位置,位移量可正、可负,相对于指令地址而言,操作数地址可以比指令地址大也可以比指令地址小。

由于大多数访存的位置都相对靠近正在执行的指令位置,则使用相对寻址可节省指令中的地址码位数,而且采用相对寻址方式编制程序,不须指定绝对地址,只须确定程序内部的相对距离,因而可以使用浮动地址,给编程带来了方便。例如,现行指令地址为 1000H,位移量为 03H,则操作数在 1003H 之中。如果在程序重定位时这段程序被安放在另一存储区域,现行指令地址改为 A000H,则操作数地址也相应改为 A003H,操作数与指令之间仍然相距 3 个单元。

位移量的确定是一个比较复杂的问题,特别是对于变字长指令更加麻烦。有些计算机是以当前指令地址为基准的,有些计算机则是以下条指令地址为基准的。这是因为有的机器是在当前指令执行完时,才将 PC 的内容加 1(或加增量);而有的机器是在取出当前指令后立即将 PC 的内容加 1(或加增量),使之变成下条指令的地址。所以位移量计算时考虑或不考虑 PC 值的更新(地址增量)都是可以的。一般来说,在取指令阶段就更新 PC 值的机器比较多,如果习题中特别指出了指令的长度和编址单位,位移量计算时应考虑 PC 值的更新问题。

不过在实际应用时,位移量是由汇编程序自动形成的,程序员在用汇编语言编写程序

时,只须写出转移目的地的标号即可。

12. 存储器堆栈的操作

存储器堆栈的栈底固定,栈顶浮动,需要一个专门的硬件寄存器作为堆栈栈顶指针(SP)。在主教材中提到了自底向上生成和自顶向下生成的两种存储器堆栈,它们在进、出栈时栈指针的修改是不同的。

对于自底向上生成的堆栈,进栈时先修改栈指针($((SP)-1 \rightarrow SP)$),然后再压入数据;出栈时,先将数据弹出,然后再修改栈指针($((SP)+1 \rightarrow SP)$)。对于自顶向下生成的堆栈,进栈时,先修改栈指针($((SP)+1 \rightarrow SP)$),然后再压入数据;出栈时,先将数据弹出,然后再修改栈指针($((SP)-1 \rightarrow SP)$)。

在讲解堆栈的进、出栈概念时,不要忘记提醒学生注意一个条件,这就是:通常栈指针始终指向栈顶的满单元。如果栈指针不是指向栈顶的满单元,而是指向栈顶的空单元,则进栈时,应先将数据压入堆栈,再将 SP 的内容自动增/减量;出栈时,应先将 SP 的内容自动增/减量,再将堆栈中的数据弹出。

对于堆栈计算机,除了上述的进栈和出栈指令以外,往往还包括一些运算类指令。表 3-1 列出了一些面向堆栈的操作指令,其中的一元操作和二元操作是为了与堆栈交换数据而设计的,本身并没有运算功能。堆栈数据运算的重要特点是指令只须指出进行什么样的操作,而无须指出操作数地址,因为地址就是栈顶,所以堆栈运算指令是零地址指令。

表 3-1 面向堆栈的操作指令

进栈指令(PUSH)	在栈顶增加一个新元素
出栈指令(POP)	从栈顶取走一个元素
一元操作	对栈顶元素进行操作以后,用结果替换栈顶元素
二元操作	对栈顶的两个元素进行操作以后,用结果替换栈顶元素

前面所描述的零地址指令程序中的运算类指令都属于二元操作指令,一条零地址加法指令的执行过程如图 3-2 所示。

需要告诉学生的是,在“计算机组成原理”课程中并没有对堆栈中的数据位数做任何限制,所以在堆栈操作指令中所有的修改栈指针操作都是 ± 1 ,也就是一个元素放在堆栈的一个单元中。对于 80x86 来说,存储器按字节编址,堆栈数据为 16 位或 32 位,此时在堆栈操作指令中所有的修改栈指针操作就必须是 ± 2 或 ± 4 ,也就是说一个元素放在堆栈的 2 个或 4 个单元中。

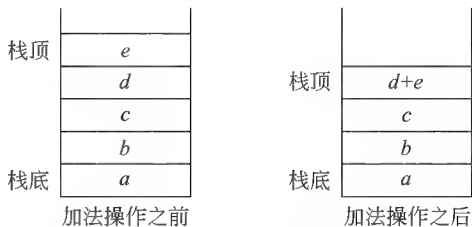


图 3-2 零地址加法指令的执行

13. 返回指令的地址字段

子程序的最后一条指令一定是返回指令。返回指令是一地址指令还是零地址指令决定于返回地址存放的位置。通常返回地址保存在堆栈中,所以返回指令 RET 后面无需任何显地址,直接从堆栈的栈顶单元就可以取出返回指令。但是对于没有堆栈的计算机,因为返回地址被保存在其他地方,所以返回指令必须是一条一地址指令,且需要通过间接寻址才能得到返回地址。

14. 输入输出指令的设置

输入输出指令是指令系统中必不可少的指令。输入与输出都以主机为参考点,由外部将信息送入主机,称为输入;由主机将信息送至外设,称为输出。主机方面的数据发送地或接收地,既可以是 CPU 中的寄存器,也可以是主存储器。外设方面通过 I/O 接口与系统总线相连,从而实现与主机的信息传送,所以外设方面的数据发送地或接收地一般是 I/O 接口中的寄存器。

各种不同计算机设置的输入输出指令差别很大,主要有下面两种类型。

(1) 采用专门的 I/O 指令

外设寄存器与主存单元分别独立编址的计算机的指令系统中都设置有专门的 I/O 指令。I/O 指令的操作码字段明确地规定某种输入输出操作,地址码字段分别给出 CPU 寄存器编号和 I/O 端口地址。例如输入指令 IN R_0, n ;其操作含义是:将端口地址为 n 的 I/O 接口寄存器的内容输入 CPU 的 R_0 寄存器中。

(2) 采用通用的数据传送指令实现 I/O 操作

外设寄存器和主存单元统一编址的计算机的指令系统中不设置专门的 I/O 指令,而采用通用的数据传送指令实现输入输出操作。如果传送指令的源地址是 CPU 寄存器,而目的地址是接口寄存器,则这条传送指令就是一条输出指令。反之,如果传送指令的源地址是接口寄存器,目的地址是 CPU 中的寄存器,则是一条输入指令。这种 I/O 指令是隐含在传送指令之中的,所以又称为隐式 I/O 指令。

例 3-6 假设某外设接口中有 3 个寄存器,它们通过数据总线与 CPU 相连,其总线地址如下:

数据寄存器	FF00H
命令字寄存器	FF01H
状态字寄存器	FF02H

下列 4 条指令完成的操作分别是什么?

- (1) MOV FF01H, R_0
- (2) MOV R_1 , FF02H
- (3) MOV R_2 , FF00H
- (4) MOV FF00H, R_3

解：(1) 执行 MOV FF01H, R₀ ;将 R₀ 内容输出到接口的命令字寄存器, R₀ 中的内容是命令字(8 位)。

(2) 执行 MOV R₁, FF02H ;将接口的状态字(8 位)输入到 CPU 的 R₁ 寄存器, 供分析判断。

(3) 执行 MOV R₂, FF00H ;将接口数据寄存器的内容输入到 CPU 的 R₂ 寄存器。

(4) 执行 MOV FF00H, R₃ ;将 R₃ 中的数据输出到接口的数据寄存器, 再传送给外设。

3.3 相关知识介绍

1. 操作码优化法 —— Huffman 编码

研究操作码的优化问题, 就是要在足够表达全部指令的前提下, 使操作码字段占用的位数最少。

最优化的编码方式是 Huffman 编码法, 它编码的原则是: 对使用频度(指在程序中出现的概率)较高的指令, 分配较短的操作码字段; 对使用频度较低的指令, 分配较长的操作码字段。每条指令在程序中使用的频度, 一般可通过大量的典型程序进行统计而求得。如果指

令系统共有 n 条指令, 则其平均码长 $\sum_{i=1}^n P_i L_i$ 较之等长操作码编码的平均码长 $\log_2 n$ 短。

在进行操作码优化时, 先构造 Huffman 树。构造 Huffman 树的方法是: 将所有指令的使用频度由小到大排序, 每次选择其中两个频度最小的结点合并成一个频度是它们两者之和的新结点, 再放到余下的结点之中, 继续找出两个频度最小的结点再结合, 直至全部频度结合完毕形成根结点为止。然后, 对每个结点的两个分支分别用二进制的 0 和 1 来标识。这样, 从根结点出发到不同频度的叶结点间所经过的 0、1 代码就是该指令的 Huffman 编码。

Huffman 编码的具体码值不惟一, 但平均码长肯定是惟一的, 而且是平均码长最短的二进制位编码。

例 3-7 某机 14 条指令的使用频度分别为: 0.01, 0.15, 0.12, 0.03, 0.02, 0.04, 0.02, 0.04, 0.01, 0.13, 0.15, 0.14, 0.11, 0.03。分别求出等长码、Huffman 码的操作码平均码长。

解: 等长操作码就是不管指令的使用频度如何, 都用同样长度的二进制码位数来对指令操作码编码。现指令系统中的指令条数为 14 条, 故平均码长等于 $\log_2 14 \approx 4$ 位。

Huffman 编码是用 Huffman 树得到的。构造 Huffman 树的方法首先是将所有指令的使用频度由小到大排序, 有:

0.01, 0.01, 0.02, 0.02, 0.03, 0.03, 0.04, 0.04, 0.11, 0.12, 0.13, 0.14, 0.15, 0.15。

然后每次选择其中两个频度最小的结点合并成一个新结点, 再放到余下的结点之中, 继续找出两个频度最小的结点再结合, 直至全部频度结合完毕形成根结点为止。得到的 Huffman 树如图 3-3 所示。平均码长为:

$$\begin{aligned} \sum_{i=1}^{14} P_i L_i &= 3 \times (0.15 + 0.15 + 0.14 + 0.13 + 0.12 + 0.11) \\ &\quad + 4 \times (0.04) + 5 \times (0.04 + 0.03 + 0.03 + 0.02 + 0.02) \\ &\quad + 6 \times (0.01 + 0.01) = 3.38 \text{ 位} \end{aligned}$$

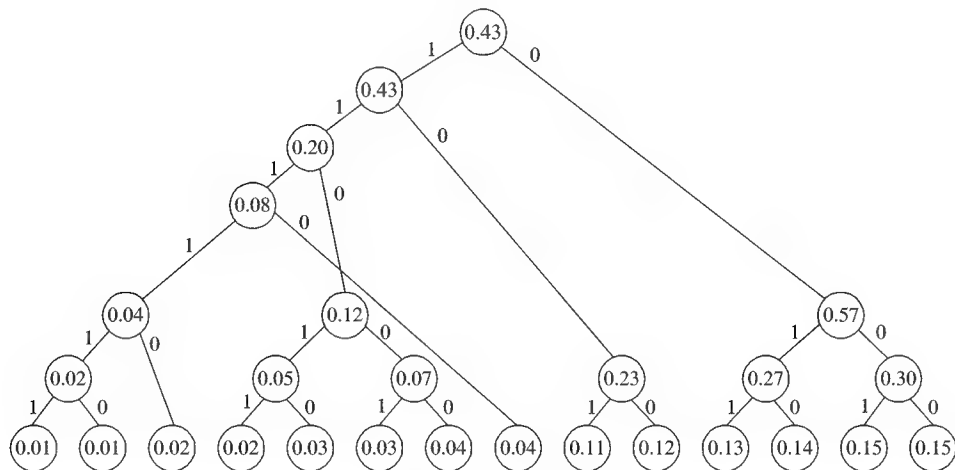


图 3-3 Huffman 树

表 3-2 列出了等长编码、Huffman 编码的例子。

表 3-2 指令操作码的几种编码与平均码长

指令 I_i	使用频度 P_i	Huffman 编码	等长编码
I_1	0.15	000	0000
I_2	0.15	001	0001
I_3	0.14	010	0010
I_4	0.13	011	0011
I_5	0.12	100	0100
I_6	0.11	101	0101
I_7	0.04	1110	0110
I_8	0.04	11000	0111
I_9	0.03	11001	1000
I_{10}	0.03	11010	1001
I_{11}	0.02	11011	1010
I_{12}	0.02	11110	1011
I_{13}	0.01	111110	1100
I_{14}	0.01	111111	1101
$\sum P_i L_i$		3.38	4.00

2. 操作码优化法 —— 扩展操作码

Huffman 编码虽然可以使信息的冗余量最小,但是形成的操作码很不规整,每条指令的操作码字段的位数都可能不同,既不便于译码,也不适于实际应用。一种实际可行的优化编码方法就是扩展操作码法,它是介于定长编码和 Huffman 编码之间的编码方式,操作码字段的位数既不是固定的,又不是任意的,而是有限的几种码长。这种扩展操作码法仍然采用了 Huffman 编码的思想,即使用频度高的指令操作码字段短,使用频度低的指令操作码字段长,从而使操作码字段的平均长度缩短,以降低信息的冗余量。应当注意,这里提到的扩展操作码法与主教材中所讨论的扩展操作码的含义有所不同,这里是从指令的使用频度出发,而不是从地址码的个数出发来决定操作码字段位数的,此时对指令的长度是没有限制的。

扩展操作码有等长扩展和不等长扩展两种方式。等长扩展是指每次扩展的操作码的位数相同,例如 4-8-12 扩展法、3-6-9 扩展法、4-6-8 扩展法均属于等长扩展;不等长扩展是指每次扩展的操作码的位数不相同,例如 4-6-10 扩展法、3-6-10 扩展法均属于不等长扩展。

在实际的计算机中,有很多都采用了扩展操作码法。其中:比较成功的当属 B-1700 机,该机指令的操作码字段有 4 位、6 位、10 位 3 种长度,高 4 位编码的 16 种组合中的 10 种用来表示 10 条使用频度最高的指令,其余 5 种组合用作扩展标志,用以指明操作码字段为 6 位长的 20 条指令(每个标志指明 4 条指令);最后一种组合也用来作扩展标志,用以指明操作码字段是 10 位长的 64 条指令,如图 3-4 所示。这种 4-6-10 的扩展操作码方案使整个指令系统所有指令的操作码字段平均位数很接近于 Huffman 编码法。

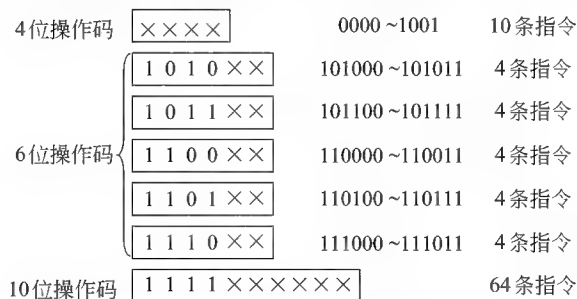


图 3-4 B 1700 机的指令格式

为了便于指令译码,最好让操作码字段等长地扩展,如 4-8-12 等。以 4-8-12 扩展为例,编码方案也很多,典型的两种方案是 15/15/15...法和 8/64/512 法。

图 3-5(a)为 15/15/15...法示意图,这种方法比较简单。图 3-5(b)为 8/64/512 示意图,头 4 位的 0×××表示最常用的 8 条指令;而后操作码字段扩展成两个 4 位,用 1×××0×××的 64 种组合表示 64 条指令;最后操作码字段扩展成 3 个 4 位,用 1×××1×××0×××的 512 种组合表示 512 条指令。

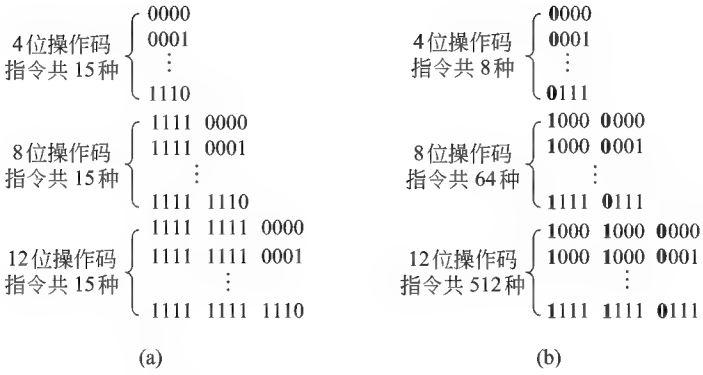


图 3-5 两种典型的扩展操作码编码方案

具体使用哪种编码方案取决于系统中指令使用频度的分布情况。如果头 15 条指令的使用频度比较大,另 15 条指令次之,其余指令使用频度很小,则宜选用 15/15/15…法。如果头 8 条指令的使用频度较大,而之后的 64 条指令的使用频度也不是过小,则宜选用 8/64/512 法。

例 3-7 中的 14 条指令若采用只有两种码长的扩展操作码,将 14 条指令按使用频度大小分组,让使用频度较高的 6 条指令用 3 位操作码编码。留下两个编码作为长码的扩展标志,进而用 5 位操作码就可以各扩展出 4 条使用频度较低的指令,这样,共有 8 条使用频度较低的指令,表 3-3 列出两种码长的扩展操作码编码。

表 3 3 两种码长的扩展操作码编码

指令 I _i	使用频度 P _i	扩展操作码编码
I ₁	0.15	000
I ₂	0.15	001
I ₃	0.14	010
I ₄	0.13	011
I ₅	0.12	100
I ₆	0.11	101
I ₇	0.04	11000
I ₈	0.04	11001
I ₉	0.03	11010
I ₁₀	0.03	11011
I ₁₁	0.02	11100
I ₁₂	0.02	11101
I ₁₃	0.01	11110
I ₁₄	0.01	11111

平均码长为：

$$\sum_{i=1}^{14} P_i L_i = 3 \times 0.80 + 5 \times 0.20 = 3.4 \text{ 位}$$

3. 面向不同对象的寻址方式

多数计算机都将主存、寄存器、堆栈分类编址,分别有面向主存、寄存器和堆栈的寻址方式。面向寄存器的寻址主要访问寄存器,少量访问主存。面向堆栈的寻址主要访问堆栈,少量访问主存和寄存器。面向主存的寻址主要访问主存,少量访问寄存器。3种面向的寻址各有特点。例如,面向堆栈的寻址有利于减轻对高级语言编译的负担,不用考虑寄存器的优化分配和使用,有利于支持子程序嵌套、递归调用时的参数、返回地址及现场等的保存和恢复。堆栈寻址可省去许多地址字段,省程序空间,存储效率高,免去了复杂的地址计算。但面向寄存器的寻址不用访存,速度比面向堆栈的快得多,因此,对向量、矩阵的运算用面向寄存器的寻址更好。

4. 变址寻址和间接寻址的比较

对于数组运算,通常要用一个循环程序对数组中的各个元素进行操作,这时必须通过修改操作数的地址才能实现。间接寻址方式与变址寻址方式的设计目标都是为了解决操作数地址的修改问题。它们都可以在程序执行过程中对操作数的地址进行修改,而不必去修改程序中的指令本身。

例 3-8 一个由 N 个元素组成的数组,已经存放在主存的连续存储单元中,现要把它搬到主存的另一个连续的存储单元中,源数组的起始地址为 AS ,目标数组的起始地址为 AD ,不必考虑可能出现的存储单元的重叠问题。首先,用间接寻址方式编写程序:

START:	MOVE ASR,ASI	;保存源数组的起始地址
	MOVE ADR,ADI	;保存目标数组的起始地址
	MOVE NUM,CNT	;保存数据的个数
LOOP:	MOVE@ ASI,@ ADI	;用间接寻址方式传送数据
	INC ASI	;源数组的地址增量
	INC ADI	;目标数组的地址增量
	DEC CNT	;个数减 1
	BGT LOOP	;数据是否传送完毕
	HALT	;停机
ASR:	AS	;源数组的起始地址
ADR:	AD	;目标数组的起始地址
NUM:	N	;需要传送的数据个数
ASI:	0	;当前正在传送的源数组地址
ADI:	0	;当前正在传送的目标数组地址
CNT:	0	;剩余数据的个数

为了程序具有再入性,前3条指令是必需的。

然后,用变址寻址方式编写程序:

```
START:  MOVE 0,X           ;变址寄存器初值为0
        MOVE NUM,CNT      ;保存数据的个数
LOOP:   MOVE AS+X,AD+X    ;用变址寻址方式传送数据
        INC X             ;增量变址寄存器
        DEC CNT           ;个数减1
        BGT LOOP         ;数据是否传送完毕
        HALT              ;停机
NUM:    N                  ;需要传送的数据个数
CNT:    0                  ;剩余数据的个数
```

比较以上两个程序,可以很明显地看出,采用变址寻址方式编写的程序简单、易读。

注:以上两段程序中的所有二地址指令均用 A_1 表示源操作数地址, A_2 表示目的操作数地址。

5. 程序在主存中的定位技术

当程序装入物理主存时,需要进行逻辑地址空间到物理地址空间的转换,即进行程序的定位。程序定位所采用的技术有静态再定位和动态再定位两种。

静态再定位是在目的程序装入主存时,通过调用系统配备的装入程序,运行此装入程序把目的程序的逻辑地址用软的方法逐一修改成物理地址。程序执行时,物理地址就不能再改变了。静态再定位不利于多道程序的运行环境,也不利于程序的可重入,同时不利于重叠、流水技术的使用。

动态再定位是指在执行每条指令时才形成访存物理地址。常用基址寻址方式实现逻辑地址到物理地址的转换,如图3-6所示。在程序装入主存时,只需要将装入主存的起始地址存入该道程序的基址寄存器中即可,指令的地址字段不做修改。程序在执行过程中,不断将逻辑地址经地址加法器加上基址寄存器中的基址,才形成物理地址。

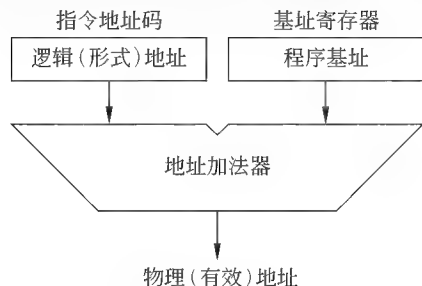


图3-6 基址寻址

6. 缩短指令中地址码长度的方法

计算机系统的主存容量通常都很大,而且会越来越大。另外,由于普遍采用了虚拟存储器结构,指令中给出的地址码是一个虚拟地址,其长度可能比实际主存的容量所要求的长度还要长得多。对于多地址结构的指令系统而言,如此长的地址码是无法容忍的。因此,如何缩短地址码的长度,是设计指令系统必须要考虑的一个问题。

由于在一般计算机系统中虚拟地址空间的大小是确定的,因此,缩短地址码长度的根本目的是要用一个比较短区域码表示一个比较大的虚拟地址空间,同时也要求有比较灵活有效的寻址方式。

缩短地址码长度的方法很多,如以下是常用方法。

① 用寄存器间接寻址方式缩短地址码长度是最有效的方法。由于寄存器的数量比较少,通常表示一个寄存器的地址只需要很少几位,而一个寄存器的字长是可以放下一个逻辑地址。例如,有8个用于间接寻址的寄存器,每个寄存器的长度是32位,这样,用一个3位的地址码就能表示一个32位的逻辑地址,再加上寻址方式等信息,一个地址码的长度也不超过10位。

② 用间接寻址方式缩短地址码长度。在主存的低端开辟出一个专门用来存放地址的区域,由于表示存储器低端部分的地址所需要的地址码长度可以很短,而一个存储字(一次访问存储器所能获得的数据)的长度通常与一个逻辑地址码的长度相当。例如,在主存最低端的1KB单元是一个用来存放地址码的区域,如果主存是按字节编址的,并且,一个存储器字的长度为32位,那么,在指令中只要用10位长度就可以表示一个32位长的逻辑地址,即使再加上寻址方式等信息,一个地址码的长度也只有十多位。

③ 用基址寻址方式缩短地址码长度。由于程序的局部性,在基址寻址中使用的地址位移量可以比较短。通常可以把比较长的基地址放在基址寄存器中,在指令的地址码中只需给出比较短的地址位移量。

7. 存储器堆栈组织

目前多数计算机的堆栈是在主存中开辟一个堆栈区,为了避免堆栈区与其他存储区混淆,堆栈除需要有栈顶指针寄存器SP以外,还应当设置堆栈上下界标志寄存器。图3-7(a)表示存储器堆栈的组织结构,其中B(Bottom)寄存器为堆栈下界(栈底)指针,L(Limit)寄存器为堆栈上界(栈顶)指针。对于自底向上生成的堆栈,在进栈时, $SP > L$ 称为堆栈“上溢”,这是不允许的;在出栈时, $SP < B$ 称为堆栈“下溢”,也是不允许的。因此,堆栈操作时,不但SP要修改,还需要判界。若采用硬件方法实现修改栈指针 $SP \leftarrow SP \pm 1$,以及判界操作,将会大大加速堆栈操作速度。

为了进一步提高堆栈操作速度,还可以将栈顶部的两个元素取到寄存器中,如图3-7(b)所示。指令对栈顶和次栈顶的数据操作可直接在寄存器中进行,此时,栈顶指针SP指向第三个栈元素。有的堆栈计算机,如HP3000,将堆栈的头4个元素存于寄存器中。这种

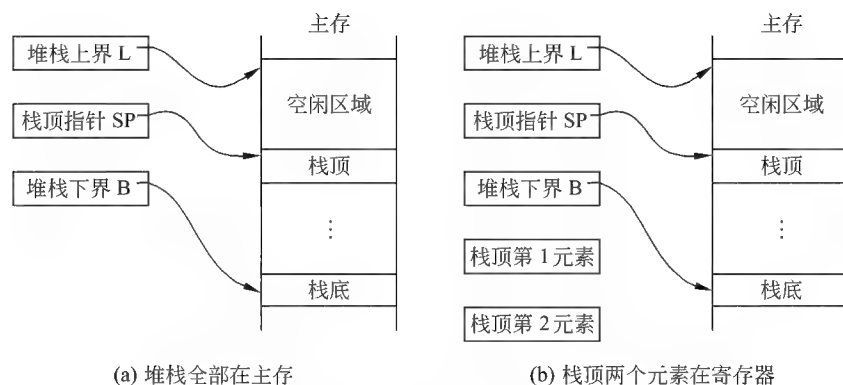


图 3-7 典型的堆栈组织

设置栈顶元素寄存器的硬件结构在面向堆栈运算的计算机中是普遍采用的。

综合起来,硬件对堆栈提供的支持有:栈顶指针 SP 及其操作时的修改,堆栈的上、下界寄存器及其判界,以及栈顶元素寄存器 3 种。

8. 其他程序控制类指令

程序控制类指令是指令系统中必不可少的一类指令,不同计算机的程序控制类指令的数量会有所不同,除过去讨论过的转移指令、子程序调用与返回指令以外,有些计算机中还包括跳越指令、循环控制指令和程序自中断指令。

(1) 跳越指令

这是一种特殊的转移指令,指令中隐含了一个地址,即下一条指令的地址。所以它只跳越过一条指令。

由于跳越指令功能简单且不需要目标地址字段,所以可以让它顺带完成其他一些功能,“加 1—判 0—跳越 (ISZ)”指令就是一个典型的例子。它实际上是一条条件跳越指令,可以用来实现迭代循环,如以下程序段:

```

301                                ;循环开始
...                                ...
309  ISZ  Ri                      ;加 1—判 0—跳越,结束循环
310  BR   301                      ;继续循环
311

```

其中, R_i 的初值为循环次数的负数补码,在循环的末尾,有 ISZ 指令把 R_i 加 1,并判是否为 0。若不为 0,则执行 BR 指令,转移到循环的开始;否则,跳过 BR 指令,退出循环。

(2) 循环控制指令

有了条件转移指令就可以实现循环程序设计,但有的计算机为了提高指令系统的有

效性,加快对循环程序的设计,还专门设置了循环控制指令。这种指令实际上是增强型的条件转移指令,它包括对循环变量的修改和结束循环条件的判断,集运算、测试和条件转移于一体,是一种具有复合功能的指令。例如,Pentium 指令系统中的 LOOPZ 指令,其功能为循环,同时循环变量(存于通用寄存器 ECX 中)自增或自减,直到循环变量值为 0,循环结束。

(3) 程序自中断指令

通常,中断是由计算机内部突发事件或外部设备的请求而随机产生的,但在有些计算机中,为了在程序调试中设置断点或实现系统调用功能,设置了专门的自中断指令。由于这些指令是由软件驱动的,所以又称为软中断,如 Intel 80x86 中的中断指令 INT n 。INT n 指令可暂停其后继指令的执行,转去执行相应的中断服务程序,但指令中不直接给出中断服务程序入口地址,而只给出中断类型码 n ,CPU 根据中断类型码可从中断向量表中找到中断服务程序的入口地址。

9. 对指令系统的基本要求

指令系统的性能如何,决定了计算机的基本功能,因而指令系统的设计是计算机系统设计中的一个核心问题。它不仅与计算机的硬件结构紧密相关,而且直接关系到用户的使用需求。不同类型计算机都有各具特色的指令系统。由于计算机性能、机器结构和特点、使用环境等要求不同,指令系统间的差异是很大的。同时,随着计算机的迅速发展,对计算机性能要求越来越高,因此,企图给计算机指令系统确定一个统一的衡量标准是很困难的。只能讨论在一般情况下,一个完善的指令系统应满足的一些基本要求,这就是:指令系统的完备性、有效性、规整性和兼容性。

(1) 完备性

完备性是指在一个有限可用的存储空间,对于任何可解的问题,在编制计算程序时,指令系统所提供的指令足够使用。完备性要求指令系统的指令丰富、功能齐全和使用方便。完备性只是一个原则性的要求,很难确定一个完备性的标准。一般来说,一个完备的指令系统应至少包括数据传送类指令、运算类指令、程序控制类指令、输入输出类指令等几种类型的指令。为了使程序能高效运行和便于硬件实现,实际计算机指令系统中实现的指令远远超过了基本完备性的要求。

(2) 有效性

有效性是指该指令系统所编制的程序能高效率的运行。所谓高效率主要表现在执行速度快、占用存储空间小两个方面。有效性针对整个指令系统而言,是一个很复杂的问题,也难以确定一个统一的标准。它与完备性是密切相关的,一个功能齐全的指令系统必定会有高的有效性。如目前许多计算机中增设的数据转换指令、字符串操作指令等,对于数据处理就会有较高的有效性。一般来说,一个功能更强、更完善的指令系统,必定有更好的有效性。

(3) 规整性

规整性包括指令的对称性、均齐性、指令格式和数据格式的一致性等特性。

对称性是指在指令系统中所有的数据存储单元(如寄存器、主存单元等)被同等对待,所有的指令都可以使用各种寻址方式。这种操作的对称性对于提高软件效率和使用方便是很有利的。如传送指令既有 $A \leftarrow B$, 也有 $B \leftarrow A$; 加法指令既有 $A \leftarrow (A + B)$, 也有 $B \leftarrow (A + B)$, 等等。

均齐性是指同一种操作性质的指令, 可以支持各种不同数据类型和不同字长的运算。例如, 加法指令能支持不同数据类型(如定点数、浮点数、十进制数等)和不同字长(如字节、字和双字, 甚至四倍字)的运算。操作的均齐性可使汇编程序与编译程序无需依赖数据类型而选用指令, 可以缩短程序空间和加快程序执行速度。

指令格式与数据格式一致性是指指令长度与数据长度有一定的关系, 以利于存取和处理。指令长度一般取字节的整倍数, 数据长度则取字节的 1、2、4 或 8 倍不等。

(4) 兼容性

不同的机器结构, 指令系统不同。但同一系列的机型则具有相同的基本结构和共同的基本指令集, 故指令系统是兼容的。由于系列机中不同机型推出的时间先后不同, 结构和性能上存在着差异, 不可能做到全部软件兼容。通常在高档机上可以运行低档机的软件, 而在低档机上则不一定能运行高档机的软件, 因此称为“向上兼容”。

3.4 教材习题解答

1. 指令长度和机器字长有什么关系? 半字长指令、单字长指令、双字长指令分别表示什么意思?

解: 指令长度与机器字长没有固定的关系, 指令长度可以等于机器字长, 也可以大于或小于机器字长。通常, 把指令长度等于机器字长的指令称为单字长指令; 指令长度等于半个机器字长的指令称为半字长指令; 指令长度等于两个机器字长的指令称为双字长指令。

2. 零地址指令的操作数来自哪里? 一地址指令中, 另一个操作数的地址通常可采用什么寻址方式获得? 各举一例说明。

解: 双操作数的零地址指令的操作数来自堆栈的栈顶和次栈顶。双操作数的一地址指令的另一个操作数通常可采用隐含寻址方式获得, 即将另一操作数预先存放在累加器中。例如, 前述零地址和一地址的加法指令。

3. 某机为定长指令字结构, 指令长度 16 位; 每个操作数的地址码长 6 位, 指令分为无操作数、单操作数和双操作数三类。若双操作数指令已有 K 种, 无操作数指令已有 L 种, 问单操作数指令最多可能有多少种? 上述三类指令各自允许的最大指令条数是多少?

解: $X = (2^4 - K) \times 2^6 - \frac{L}{2^6} k$

双操作数指令的最大指令数: $2^4 - 1$ 。

单操作数指令的最大指令数: $15 \times 2^6 - 1$ (假设双操作数指令仅 1 条, 为无操作数指令留出 1 个扩展窗口)。

无操作数指令的最大指令数: $2^{16} - 2^{12} - 2^6$ 。其中 2^{12} 为表示某条二地址指令占用的编码数, 2^6 为表示某条单地址指令占用的编码数。此时双操作数和单操作数指令各仅有 1 条。

4. 设某机为定长指令字结构, 指令长度 12 位, 每个地址码占 3 位, 试提出一种分配方案, 使该指令系统包含: 4 条三地址指令, 8 条二地址指令, 180 条单地址指令。

解: 4 条三地址指令

000 XXX YYY ZZZ

...

011 XXX YYY ZZZ

8 条二地址指令

100 000 XXX YYY

...

100 111 XXX YYY

180 条单地址指令

101 000 000 XXX

...

111 110 011 XXX

5. 指令格式同上题, 能否构成:

三地址指令 4 条, 单地址指令 255 条, 零地址指令 64 条? 为什么?

解: 三地址指令 4 条

000 XXX YYY ZZZ

...

011 XXX YYY ZZZ

单地址指令 255 条

100 000 000 XXX

...

111 111 110 YYY

只能再扩展出零地址指令 8 条, 所以不能构成这样的指令系统。

6. 指令中地址码的位数与直接访问的主存容量和最小寻址单位有什么关系?

解：主存容量越大,所需的地址码位数就越长。对于相同容量来说,最小寻址单位越小,地址码的位数就越长。

7. 试比较间接寻址和寄存器间址。

解：间接寻址方式的有效地址在主存中,操作数也在主存中;寄存器间址方式的有效地址在寄存器中,操作数在主存中。所以间接寻址比较慢。

8. 试比较基址寻址和变址寻址。

解：基址寻址和变址寻址在形成有效地址时所用的算法是相同的,但是它们两者实际上是有区别的。一般来说,变址寻址中变址寄存器提供修改量(可变的),而指令中提供基准值(固定的);基址寻址中基址寄存器提供基准值(固定的),而指令中提供位移量(可变的)。这两种寻址方式应用的场合也不同,变址寻址是面向用户的,用于访问字符串、向量和数组等成批数据;而基址寻址面向系统,主要用于逻辑地址和物理地址的变换,用以解决程序在主存中的再定位和扩大寻址空间等问题。在某些大型机中,基址寄存器只能由特权指令来管理,用户指令无权操作和修改。

9. 某机字长为 16 位,主存容量为 64K 字,采用单字长单地址指令,共有 50 条指令。若有直接寻址、间接寻址、变址寻址、相对寻址四种寻址方式,试设计其指令格式。

解：操作码 6 位,寻址方式 2 位,地址码 8 位。

10. 某机字长为 16 位,主存容量为 64K 字,指令格式为单字长单地址,共有 64 条指令。试说明:

(1) 若只采用直接寻址方式,指令能访问多少主存单元?

(2) 为扩充指令的寻址范围,可采用直接/间接寻址方式,若只增加一位直接/间接标志,指令可寻址范围为多少? 指令直接寻址的范围为多少?

(3) 采用页面寻址方式,若只增加一位 Z/C(零页/现行页)标志,指令寻址范围为多少? 指令直接寻址范围为多少?

(4) 采用(2)、(3)两种方式结合,指令的寻址范围为多少? 指令直接寻址范围为多少?

解：因为计算机中共有 64 条指令,所以操作码占 6 位,其余部分为地址码或标志位。

(1) 若只采用直接寻址方式,地址码部分为 10 位,指令能访问的主存单元数为 $2^{10} = 1\text{K}$ 字。

(2) 若采用直接/间接寻址方式,将增加了一位直接/间接标志,地址码部分为 9 位,指令直接寻址的范围为 $2^9 = 0.5\text{K}$ 字,指令可寻址范围为整个主存空间 $2^{16} = 64\text{K}$ 字。

(3) 若采用页面寻址方式,将增加一位 Z/C(零页/现行页)标志,所以指令直接寻址范围仍为 $2^9 = 0.5\text{K}$ 字,指令寻址范围仍为 $2^{16} = 64\text{K}$ 字。

(4) 此时将需要 Z 和 C 两个标志位,所以指令直接寻址范围为 $2^8 = 0.25\text{K}$ 字,指令的可寻址范围仍为 $2^{16} = 64\text{K}$ 字。

11. 设某机字长 32 位, CPU 有 32 个 32 位的通用寄存器, 设计一个能容纳 64 种操作的单字长指令系统。

(1) 如果是存储器间接寻址方式的寄存器—存储器型指令, 能直接寻址的最大主存空间是多少?

(2) 如果采用通用寄存器作为基址寄存器, 能直接寻址的最大主存空间又是多少?

解: 因为计算机中共有 64 条指令, 所以操作码占 6 位; 32 个通用寄存器, 寄存器编号占 5 位; 其余部分为地址码或标志位。

(1) 如果是存储器间接寻址方式的寄存器—存储器型指令, 操作码 6 位, 寄存器编号 5 位, 间址标志 1 位, 地址码 20 位, 直接寻址的最大主存空间是 2^{20} 字。

(2) 如果采用通用寄存器作为基址寄存器, $EA = (R_b) + A$, 能直接寻址的最大主存空间是 2^{32} 字。

12. 已知某小型机字长为 16 位, 其双操作数指令的格式如下:

0	5	6	7	8	15
OP			R	A	

其中: OP 为操作码, R 为通用寄存器地址。试说明下列各种情况下能访问的最大主存区域有多少机器字?

(1) A 为立即数。

(2) A 为直接主存单元地址。

(3) A 为间接地址(非多重间址)。

(4) A 为变址寻址的形式地址, 假定变址寄存器为 R_1 (字长为 16 位)。

解: (1) 1 个机器字。

(2) 256 个机器字。

(3) 65536 个机器字。

(4) 65536 个机器字。

13. 计算下列 4 条指令的有效地址(指令长度为 16 位)。

(1) 000000Q

(2) 100000Q

(3) 170710Q

(4) 012305Q

假定: 上述 4 条指令均用八进制书写, 指令的最左边是一位间址指示位@ (@=0, 直接寻址; @=1, 间接寻址), 且具有多重间访功能; 指令的最右边两位为形式地址; 主存容量 2^{15} 单元, 表 3-4 为有关主存单元的内容(八进制)。

表 3-4 习题 13 的表格

地址	内容
00000	100002
00001	046710
00002	054304
00003	100000
00004	102543
00005	100001
00006	063215
00007	077710
00010	100005

解: (1) 000000Q

因为指令的最高位为 0, 故为直接寻址, $EA = A = 00000Q$ 。

(2) 100000Q

因为指令的最高位为 1, 故指令为间接寻址。

(00000) = 100002, 最高位仍为 1, 继续间接寻址。

(00002) = 054304, 其最高位为 0, 表示已找到有效地址, $EA = 54304Q$ 。

(3) 170710Q

因为指令的最高位为 1, 故指令为间接寻址。

(00010) = 100005, 最高位仍为 1, 继续间接寻址。

(00005) = 100001, 最高位仍为 1, 继续间接寻址。

(00001) = 046710, 其最高位为 0, 表示已找到有效地址, $EA = 46710Q$ 。

(4) 012305Q

因为指令的最高位为 0, 故为直接寻址, $EA = A = 00005Q$ 。

14. 假定某机的指令格式如下:

11	10	9	8	7	6	5	0
@	OP	I ₁	I ₂	Z/C	A		

其中:

Bit11 = 1: 间接寻址;

Bit8 = 1: 变址寄存器 I₁ 寻址;

Bit7 = 1: 变址寄存器 I₂ 寻址;

Bit6 (零页/现行页寻址): Z/C = 0, 表示 0 页面;

Z/C = 1, 表示现行页面, 即指令所在页面。

若主存容量为 2^{12} 个存储单元, 分为 2^6 个页面, 每个页面有 2^6 个字。

设有关寄存器的内容为

(PC) = 0340Q (I₁) = 1111Q (I₂) = 0256Q

试计算下列指令的有效地址。

(1) 1046Q

(2) 2433Q

(3) 3215Q

(4) 1111Q

解: (1) 1046Q=001 000 100 110

因为 4 个标志位均为 0, 故为直接寻址, $EA = A = 0046Q$ 。

(2) 2433Q=010 100 011 011

因为 Bit8(I_1)=1, 故为变址寄存器 1 寻址, $EA = (I_1) + A = 1111 + 33 = 1144Q$ 。

(3) 3215Q=011 010 001 101

因为 Bit7(I_2)=1, 故为变址寄存器 2 寻址, $EA = (I_2) + A = 0256 + 15 = 0273Q$ 。

(4) 1111Q=001 001 001 001

因为 Bit6(Z/C)=1, 故为当前页寻址, $EA = (PC)_H // A = 03 // 11 = 0311Q$ 。

15. 假定指令格式如下:

15	12	11	10	9	8	7	0
OP	I_1	I_2	Z/C	D/I	A		

其中: D/I 为直接/间接寻址标志, D/I=0 表示直接寻址, D/I=1 表示间接寻址。其余标志位同题 14 说明。

若主存容量为 2^{16} 个存储单元, 分为 2^8 个页面, 每个页面有 2^8 个字。

设有关寄存器的内容为

(I_1)=002543Q (I_2)=063215Q (PC)=004350Q

试计算下列指令的有效地址。

(1) 152301Q

(2) 074013Q

(3) 161123Q

(4) 140011Q

解: (1) 152301Q=1 101 010 011 000 001

因为 Bit10(I_2)=1, 故为变址寄存器 2 寻址, $EA = (I_2) + A = 063215 + 301 = 063516Q$ 。

(2) 074013Q=0 111 100 000 001 011

因为 Bit11(I_1)=1, 故为变址寄存器 1 寻址, $EA = (I_1) + A = 002543 + 013 = 002556Q$ 。

(3) 161123Q=1 110 001 001 010 011

因为 Bit9(Z/C)=1, 故为当前页寻址, $EA = (PC) // A = 004123Q$ 。

(4) 140011Q=1 100 000 000 001 001

因为 4 个标志位均为 0, 故为直接寻址, $EA = A = 000011Q$ 。

16. 举例说明哪几种寻址方式除去取指令以外不访问存储器? 哪几种寻址方式除去

取指令外只需访问一次存储器？完成什么样的指令，包括取指令在内共访问 4 次存储器？

解：除去取指令以外不访问存储器：立即寻址，寄存器寻址。

除去取指令外只需访问一次存储器：直接寻址，寄存器间接寻址，变址寻址，基址寻址，相对寻址，页面寻址。

二级间接寻址包括取指令在内共访问 4 次存储器。

17. 设相对寻址的转移指令占两个字节，第一个字节是操作码，第二个字节是相对位移量，用补码表示。假设当前转移指令第一字节所在的地址为 2000H，且 CPU 每取一个字节便自动完成 $(PC)+1 \rightarrow PC$ 的操作。试问当执行 $JMP *+8$ 和 $JMP *-9$ 指令（* 为相对寻址特征）时，转移指令第二字节的内容各为多少？转移的目的地址各是什么？

解：转移指令第二字节的内容分别为：00001000(+8)，11110111(-9)

转移的目的地址分别为：200AH，1FF9H。

18. 什么叫主程序和子程序？调用子程序时还可采用哪几种方法保存返回地址？画图说明调用子程序的过程。

解：主程序就是指通常的程序，而子程序是一组可以公用的指令序列，只要知道子程序的入口地址就能调用它。

保存返回地址的方法有多种：

(1) 用子程序的第一个字单元存放返回地址。转子指令把返回地址存放在子程序的第一个字单元中，子程序从第二个字单元开始执行。返回时将第一个字单元地址作为间接地址，采用间址方式返回主程序。

(2) 用寄存器存放返回地址。转子指令先把返回地址放到某一个寄存器中，再由子程序将寄存器中的内容转移到另一个安全的地方。

(3) 用堆栈保存返回地址。

调用子程序的过程如图 3-8 所示，此时返回地址保存在堆栈中。

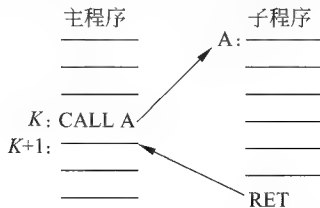


图 3-8 主程序调用子程序的过程

19. 在某些计算机中，调用子程序的方法是这样实现的：转子指令将返回地址存入子程序的第一个字单元，然后从第二个字单元开始执行子程序，请回答下列问题：

(1) 为这种方法设计一条从子程序转到主程序的返回指令。

- (2) 在这种情况下,怎么在主、子程序间进行参数的传递?
- (3) 上述方法是否可用于子程序的嵌套?
- (4) 上述方法是否可用于子程序的递归(即某个子程序自己调用自己)?
- (5) 如果改用堆栈方法,是否可实现(4)所提出的问题?

解: (1) 返回指令通常为零地址指令。返回地址保存在堆栈中,执行返回指令时自动从堆栈中弹出。而目前返回地址是保存在子程序的第一个单元中,故此时返回指令不能再是零地址指令了,而应当是一地址指令。如:

JMP	@	子程序首地址
-----	---	--------

间接寻址可找到返回地址,然后无条件转移到返回的位置。

- (2) 在这种情况下,可利用寄存器或主存单元进行主、子程序间的参数传递。
- (3) 可以用于子程序的嵌套(多重转子)。因为每个返回地址都放在调用的子程序的第一个单元中。
- (4) 不可以用于子程序的递归,因为当某个子程序自己调用自己时,子程序第一个单元的内容将被破坏。
- (5) 如果改用堆栈方法,可以实现子程序的递归,因堆栈具有后进先出的功能。

第 4 章

数值的机器运算

4.1 基本内容要求

运算器是计算机进行算术运算和逻辑运算的主要部件,运算器的逻辑结构取决于机器的指令系统、数据表示方法和运算方法等。本章主要讨论数值数据在计算机中实现算术运算和逻辑运算的方法,以及运算部件的基本结构和工作原理。

学习要求

- ◆ 了解串行加法器与并行加法器的区别
- ◆ 理解进位产生和进位传递的概念
- ◆ 掌握并行加法器不同进位方式的特点与区别
- ◆ 了解 $[-Y]_{\text{补}}$ 的含义和求 $[-Y]_{\text{补}}$ 的方法
- ◆ 掌握定点加法和减法运算方法
- ◆ 了解溢出产生的原因
- ◆ 掌握 3 种溢出检测方法的区别,特别是双符号位补码判断溢出的特点
- ◆ 掌握补码的左移、右移运算方法
- ◆ 了解常见的舍入操作方法
- ◆ 理解原码一位乘法运算方法
- ◆ 掌握补码一位乘法运算方法
- ◆ 了解补码两位乘法运算方法
- ◆ 理解原码加减交替除法运算方法
- ◆ 掌握补码加减交替除法运算方法
- ◆ 理解浮点加、减、乘、除运算的过程
- ◆ 理解 1 位十进制整数的加法运算
- ◆ 理解逻辑运算的特点
- ◆ 了解运算器的基本结构

- ◆ 掌握典型的 ALU 芯片(74181、74182)
- ◆ 了解浮点协处理器的作用

4.2 误点疑点解惑

1. 并行加法器的进位产生和传递

一个 n 位字长的并行加法器由 n 个全加器组成, n 位数据同时相加。虽然操作数的各位是同时提供的, 但低位运算所产生的进位会影响到高位的运算结果, 所以并行加法器的最长运算时间主要是由进位信号的传递时间决定的, 而每个全加器本身的求和延迟只是次要因素。很明显, 提高并行加法器速度的关键是尽量加快进位产生和传递的速度。

并行加法器中的每一个全加器都有一个从低位送来的进位和一个传送给较高位的进位, 每一位的进位表达式为:

$$C_i = A_i B_i + (A_i \oplus B_i) C_{i-1} = G_i + P_i C_{i-1}$$

G_i 称为全加器第 i 位的进位产生函数, 其逻辑含义是: 若本位两个输入均为 1, 必向高位产生进位, 与低位进位无关。 P_i 称为进位传递函数, 其逻辑含义是: 当 $P_i = 1$ 时, 若低位有进位, 本位将产生进位。

2. 并行加法器的进位传递方式和传递时间

n 位并行加法器按进位信号的传递方式, 可分为串行进位方式、并行进位方式和分组并行进位方式。

串行进位方式的每一级进位直接依赖于前一级的进位, 即进位信号是逐级形成的。假定, 一级“与门”、“或门”的延迟时间定为 t_y , 则每一级进位的延迟时间为 $2t_y$ 。在字长为 n 位的情况下, 若不考虑 G_i 、 P_i 的形成时间, 从 $C_0 \rightarrow C_n$ 的最长延迟时间为 $2nt_y$ (设 C_0 为加法器最低位的进位, C_n 为加法器最高位的进位)。串行进位速度慢, 且加法器位数越长, 进位延迟时间也越长。

并行进位方式所有各位的进位不依赖于其低位的进位, 而依赖于最低位的进位 C_0 , 各位的进位是同时产生的。这种进位方式是快速的, 若不考虑 G_i 、 P_i 的形成时间, 从 $C_0 \rightarrow C_n$ 的最长延迟时间仅为 $2t_y$, 而与字长无关。随着加法器位数的增加, 完全采用并行进位是不现实的。

真正实用的进位方式是分组先行进位方式, 分组先行进位方式又有单级和多级之分。

单级先行进位方式又称为组内并行、组间串行方式。若不考虑 G_i 、 P_i 的形成时间, 从 $C_0 \rightarrow C_n$ 的最长延迟时间为 $2mt_y$, 其中 m 为分组的组数。16 位单级先行进位加法器 (分为 4 组, 每组 4 位), 从 $C_0 \rightarrow C_{16}$ 的最长延迟时间为 $4 \times 2t_y = 8t_y$ 。

多级先行进位方式又称组内并行、组间并行方式。若不考虑 G_i 、 P_i 的形成时间, 在 16 位二级先行进位加法器中, C_0 经过 $2t_y$ 产生第 1 小组的 G_1 、 C_2 、 C_3 及所有组进位产生函数

G_i^* 和组进位传递函数 P_i^* ; 再经过 $2t_y$, 产生 C_4 、 C_8 、 C_{12} 、 C_{16} ; 最后经过 $2t_y$ 后, 才能产生第 2、3、4 小组内的 $C_5 \sim C_7$ 、 $C_9 \sim C_{11}$ 、 $C_{13} \sim C_{15}$ 。在这里一定要提醒学生注意, 从 $C_0 \rightarrow C_{16}$ 的延迟时间为 $4t_y$, 而从 $C_0 \rightarrow C_5$ 的延迟时间为 $6t_y$, 因为 C_4 和 C_{16} 是同时产生的, 如果没有 C_4 就不会产生正确的 C_5 , 此时高位的进位先于低位的进位产生, 整个加法器的最长进位延迟时间为 $6t_y$ 。

3. 补码加减运算及其实现

补码加法, 符号位参加运算, 被加数和加数直接相加, 即

$$[X+Y]_{\text{补}} = [X]_{\text{补}} + [Y]_{\text{补}}$$

如果计算机中有减法器, 则补码减法有:

$$[X-Y]_{\text{补}} = [X]_{\text{补}} - [Y]_{\text{补}}$$

而实际上, 计算机中并没有减法器, 减法也是由加法器来完成的。对于补码减法, 符号位参加运算, 被减数和减数的机器负数直接相加, 即

$$[X-Y]_{\text{补}} = [X]_{\text{补}} + [-Y]_{\text{补}}$$

从两个减法公式可以看出:

$$[-Y]_{\text{补}} = -[Y]_{\text{补}}$$

例 4-1 求证: $[-Y]_{\text{补}} = -[Y]_{\text{补}}$ 。

证明:

因为

$$[X]_{\text{补}} + [Y]_{\text{补}} = [X+Y]_{\text{补}}$$

令 $X = -Y$ 代入上式, 则有:

$$[-Y]_{\text{补}} + [Y]_{\text{补}} = [-Y+Y]_{\text{补}} = [0]_{\text{补}} = 0$$

所以

$$[-Y]_{\text{补}} = -[Y]_{\text{补}}$$

在减法运算时, 寄存器 Y 中存放着的是减数的补码形式 $[Y]_{\text{补}}$ 。已知 $[Y]_{\text{补}}$ 求 $[-Y]_{\text{补}}$ 的方法是: 将 $[Y]_{\text{补}}$ 连同符号位一起求反, 末尾加“1”。这个过程称为变补(求补), 表示为

$$[-Y]_{\text{补}} = [[Y]_{\text{补}}]_{\text{变补}}$$

初学者很容易将“某数的补码表示”与“变补”这两个概念搞混, 一定要多举几个例子。变补时, 无论 $[Y]_{\text{补}}$ 表示的真值是正数还是负数, 都要对 $[Y]_{\text{补}}$ 包括符号位一起变反(所有的二进制位一起变反), 末位加“1”。

无论是加法还是减法运算, 均用相同的逻辑电路来实现, 实现补码加减运算的逻辑电路如图 4-1 所示。

补码加减运算器的核心是一个多位的并行加法器 F , X 和 Y 是两个寄存器, $\uparrow A$ 、 B 、 C 分别是字级的与门和与或门。在两数运算之前, X 、 Y 寄存器中存放着补码表示的被操作数和操作数, 运算结束后, X 寄存器中存放着补码表示的结果。

加、减法运算的控制信号的不同之处在于, 加法时由 $Y \rightarrow F$ 信号打开与或门 B 的右侧与门, 将 $[Y]_{\text{补}}$ 的原变量送到加法器与 $[X]_{\text{补}}$ 相加, 减法时由 $Y \rightarrow F$ 信号打开与或门 B 的左

侧与1,将 $[Y]_{\text{补}}$ 的反变量送到加法器与 $[X]_{\text{补}}$ 相加,并由 $1 \rightarrow F$ 信号使加法器的最低位加“1”。

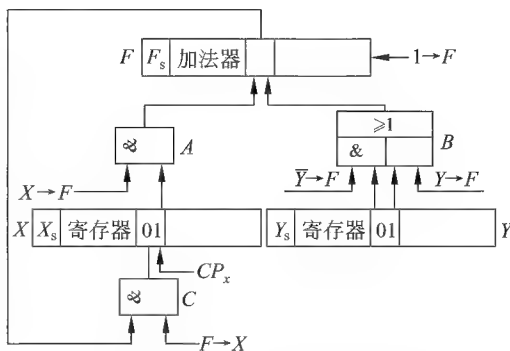


图 4 1 补码加减运算器框图

4. 补码 Booth 乘法

乘法运算需要 3 个寄存器。被乘数 $[X]_{\text{补}}$ 存放在 B 寄存器中,乘数 $[Y]_{\text{补}}$ 存放在 C 寄存器中; A 寄存器用来存放部分积与最后乘积的高位部分,它的初值为 0。运算结束后寄存器 C 中不再保留乘数,改为存放乘积的低位部分。

补码乘法运算过程中, A 、 C 两个寄存器级联起来右移。若乘数的数值位为 n 位, 共需进行 $n+1$ 次累加和 n 次右移, 最后将得到一个数值位为 $2n$ 位的乘积, 高位在 A 寄存器中, 低位在 C 寄存器中。

在 Booth 乘法运算中,学生常犯的错误的忘记在乘数的最低位之后增加一位附加位 Y_{n+1} , Y_{n+1} 的初值为 0。Booth 乘法规则中虽然每次比较两位乘数,但实际上只对一位乘数进行处理,如果不在乘数的最低位后增加 Y_{n+1} 的话,则乘数的最低位 Y_n 将不能得到处理,运算结果当然就不正确了。

5. 补码加减交替除法

除法运算也需要 3 个寄存器。被除数存放在 A 寄存器中;除数存放在 B 寄存器中;C 寄存器用来存放商,它的初值为 0。运算过程中 A 寄存器的内容将不断地发生变化,最后 A 寄存器中剩下的是扩大了若干倍的余数。

补码除法运算过程中, A、C 两个寄存器级联起来左移。若除数的数值位为 n 位, 共需进行 $n+1$ 次累加和 n 次左移, 最后得到数值位为 n 位的商和余数。

补码加减交替除法运算要比 Booth 乘法运算稍复杂一些,特别是对于够减的判断、上商规则和商符形成的理解可能会使部分学生感到困难。下面简单讨论够减的判断、上商规则和商符形成的问题,真正在计算时并不需要深究这些问题,只要按规则一步步做就可以了。

(1) 够减的判断

除法运算实际上是在做减法运算,只不过如果两数同号,则真的做减法运算;而两数异号,做减法变成做加法运算。参加运算的两个数符号任意,够减的情况如下。

① 两数同号

$$\begin{aligned} X > 0, Y > 0, X - Y > 0 \\ X < 0, Y < 0, -X - (-Y) > 0 \Rightarrow X - Y < 0 \end{aligned}$$

② 两数异号

$$\begin{aligned} X > 0, Y < 0, X - (-Y) = (X + Y) > 0 \\ X < 0, Y > 0, (-X) - Y > 0 \Rightarrow X + Y < 0 \end{aligned}$$

综合以上情况可得出下列结论:当被除数 $[X]_{\text{补}}$ (或部分余数)与除数 $[Y]_{\text{补}}$ 同号时,如果得到的新部分余数 $[r_i]_{\text{补}}$ 与除数 $[Y]_{\text{补}}$ 同号,表示够减,否则为不够减;当被除数 $[X]_{\text{补}}$ (或部分余数)与除数 $[Y]_{\text{补}}$ 异号时,如果得到的新部分余数 $[r_i]_{\text{补}}$ 与除数 $[Y]_{\text{补}}$ 异号,表示够减,否则为不够减。

(2) 上商规则

如果 $[X]_{\text{补}}$ 和 $[Y]_{\text{补}}$ 同号,则商为正数,够减时上商“1”,不够减时上商“0”;如果 $[X]_{\text{补}}$ 和 $[Y]_{\text{补}}$ 异号,则商为负数,够减时上商“0”,不够减时上商“1”。

补码的上商规则最后可归结为:部分余数 $[r_i]_{\text{补}}$ 和除数 $[Y]_{\text{补}}$ 同号,商上“1”,反之,商上“0”。

(3) 商符的确定

第一次得出的商,就是实际应得的商符。因为为了保证商是一个定点小数,必须要求 $|X| < |Y|$,所以第一次肯定不够减。当被除数与除数同号时,部分余数与除数必然异号,商上“0”,恰好与商符一致;当被除数与除数异号,部分余数与除数必然同号,商上“1”,也恰好就是商的符号。

在加减交替法除法运算中,学生常犯的错误主要有:①左移过程出错;②忘记商的最末1位应当恒置“1”。

在补码加减交替除法中采用双符号位进行运算,最左边的符号位是真符。左移时要特别注意,如:

00.1××××左移1位为01.××××0

11.0××××左移1位为10.××××0

6. 浮点加减运算中的对阶和结果规格化

浮点数的加减运算首先需要对阶,对阶的实质就是小数点对齐。对阶的原则是小阶向大阶看齐。使小阶的阶码增大,相应的尾数右移,直到两数的阶码相等为止。当尾数的基数 $r=2$ 时,每右移1位,阶码加1。

当尾数结果为00.0××××或11.1××××时,需要使尾数左移以实现规格化,这

这个过程称为左规。左规可能需要进行多次,尾数每左移1位,阶码相应减1,直至成为规格化数为止。

当尾数结果为 $10.\times\times\times\cdots\times$ 或 $01.\times\times\times\cdots\times$ 时,应将尾数右移以实现规格化,这个过程称为右规。右规最多只需要进行一次,尾数每右移1位,阶码相应加1。

在尾数左或右规以后,有可能使阶码发生溢出。若阶码用双符号位补码表示,当:

$[E_C]_{\text{补}} = 01.\times\times\times\cdots\times$, 表示上溢。此时,浮点数真正溢出,机器需停止运算,做溢出中断处理。

$[E_C]_{\text{补}} = 10.\times\times\times\cdots\times$, 表示下溢。浮点数值趋于零,机器不做溢出处理,而是当作机器零处理。

7. 浮点除法运算中的尾数调整

对尾数来说,当被除数的绝对值大于等于除数的绝对值时(即 $|M_A| \geq |M_B|$),在定点除法运算中是不允许的,但在浮点除法运算中是允许的。但由于除法规则是在 $|M_A| < |M_B|$ 的前提下推出的,为使定点除法规则在浮点除法尾数相除时也能应用,通常在尾数除法前加上尾数调整的步骤。

所谓尾数调整,是指将被除数尾数调整为小于除数的尾数,即经过调整后被除数的尾数为 M_A' ,应使 $1/2 < |M_A'| < |M_B| < 1$ 。如 $|M_A| \geq |M_B|$,则 $|M_A'| = |M_A|/2$;如 $|M_A| < |M_B|$,则 $|M_A'| = |M_A|$ 。这样做,不仅使定点除法规则可以适用于浮点除法的尾数相除,而且所得的商必为规格化的数,省去除法运算后规格化的步骤。

下面分两种情况加以证明。

① 若 $|M_A| < |M_B|$,则 M_A 不需调整, $|M_A'| = |M_A|$ 。

因为 $|M_A| < |M_B|$,故有: $|M_A'|/|M_B| < 1$ 。

对于 $|M_A'|/|M_B| \geq 1/2$,可以采用反证法证明:

假设 $|M_A'|/|M_B| = m < 1/2$,则有 $|M_A'| = m \times |M_B|$,由于 $|M_B| < 1$,故有 $|M_A| = |M_A'| = m \times |M_B| < m < 1/2$ 。这与 M_A 是规格化数矛盾。所以 $|M_A'|/|M_B| \geq 1/2$ 。

因此, $1/2 \leq |M_A'|/|M_B| < 1$ 。

② 若 $|M_A| \geq |M_B|$,则 M_A 需要调整, $|M_A'| = |M_A|/2$ 。此时必有 $|M_A'| < |M_B|$ 。

因为 $|M_A'| < |M_B|$,故有 $|M_A'|/|M_B| < 1$ 。

因为 $|M_A| \geq |M_B|$,故有 $|M_A|/|M_B| \geq 1$;

而 $|M_A'|/|M_B| = (|M_A|/2)/|M_B| = (|M_A|/|M_B|)/2 \geq 1/2$ 。

因此, $1/2 \leq |M_A'|/|M_B| < 1$ 。

综合以上两种情况可知,所得商必为规格化的数。

8. BCD 码的加法运算

BCD 码由4位二进制数表示,按二进制加法规则进行加法。十进制数的进位是10,而4位二进制数的进位是16,为此需要进行必要的十进制校正,才能使该进位正确。不

同的 BCD 码对应的十进制校正规律是不一样的,因此硬件实现也是不同的。

在主教材中已经讨论了 8421 码和余 3 码的加法规则和加法器。无论哪一种 BCD 码,都需要首先找出其校正关系,然后再根据校正关系列出校正函数,最后得到相应的 1 位加法器电路。所以说找出 BCD 码的校正关系是解决问题的关键,由于两个 1 位的十进制数(0~9)相加,其和不会超过 18,考虑低位来的进位,其和最大值是 19。校正关系表中应当列出正确的 BCD 码和校正前的二进制数,两者之间的区别就是需要校正(加或减)的数。常见的 BCD 码(8421 码、余 3 码、2421 码)的校正关系都不是很复杂,根据校正关系表找出其校正关系应该是不困难的。

9. 基本逻辑运算及其应用

逻辑运算的主要特点是:数据按位进行操作,每位均按二值布尔规则运算,各位之间无进位和借位关系,也没有溢出。

逻辑运算多用于按位或字段的处理。如用来改变某些指定位的状态;在一个字中取出一部分字段,或插入一部分新的数值;按照另外一个寄存器的内容改变现有数据等。

(1) 利用与运算实现按位测试

让屏蔽字中相应位为 1,其他位为 0,然后两个操作数相与,使需要检测的位保留原来的状态,不需要检测的位为 0。

目的操作数 <i>A</i>	1100 1010
屏蔽字 <i>B</i>	0000 1000
<i>A</i> AND <i>B</i>	0000 1000

(2) 利用与运算实现按位分离

让屏蔽字中对应于分离段的各位为 1,其他位为 0,然后两个操作数相与,以便分离出感兴趣的一段代码。

目的操作数 <i>A</i>	1100 1010
屏蔽字 <i>B</i>	0000 1111
<i>A</i> AND <i>B</i>	0000 1010

(3) 利用与运算实现按位清除

让屏蔽字中相应位为 0,其他位为 1,然后与目的操作数相与。

目的操作数 <i>A</i>	1100 1010
屏蔽字 <i>B</i>	1111 0111
<i>A</i> AND <i>B</i>	1100 0010

(4) 利用或运算实现按位设置

让屏蔽字中相应位为 1,其他位为 0,然后与目的操作数相或。

目的操作数 <i>A</i>	11001 010
屏蔽字 <i>B</i>	00000 100

$A \text{ OR } B$ 11001 110

(5) 利用异或运算实现按位修改

被处理的数中哪些位需要变反,则屏蔽字中的相应位为1,不修改的位为0,然后两操作数相异或。

目的操作数 A 1100 1010

屏蔽字 B 0000 1000

$A \text{ EOR } B$ 1100 0010

(6) 利用异或运算实现判符合

将待判定的代码与设定的代码相异或,若结果各位均为0,表示两者相同;若有一位不为0,表示两数不相同。

目的操作数 A 11001010

屏蔽字 B 11001010

$A \text{ EOR } B$ 00000000

(7) 利用与和或运算实现插入

插入是指将代码中的某些位用新的数值取代。例如,要求在 A 的前4位插入新的数值1101。首先使用与运算将 A 的前4位删除,然后再将 A 与要求插入的数值相或。

目的操作数 A 11001010

屏蔽字 B 00001111

$A \text{ AND } B$ 00001010 删除高4位

A 00001010

屏蔽字 B 11010000

$A \text{ OR } B$ 11011010 插入高4位

4.3 相关知识介绍

1. 全加器电路

全加器(FA)是最基本的加法单元,它有3个输入量:操作数 A_i 和 B_i 、低位传来的进位 C_{i-1} ,两个输出量:本位和 S_i 、向高位的进位 C_i 。

根据全加器真值表,可得到全加器的和 S_i 与进位 C_i 的逻辑表达式为:

$$S_i = A_i \oplus B_i \oplus C_{i-1}$$

$$C_i = A_i B_i + A_i C_{i-1} + B_i C_{i-1} = A_i B_i + (A_i + B_i) C_{i-1} = A_i B_i + (A_i \oplus B_i) C_{i-1}$$

图4-2 为全加器的逻辑图。

根据数字电路的知识可知,对于各种门电路,从输入信号出现到产生输出信号是有时延的,不同的门电路延迟时间不同。假设一级与非门的延迟时间为 $1t_y$,一级与或非

与、异或门的延迟时间为 $1.5t_y$, 则产生和 S_i 要经过 $3t_y$ 延时, 产生进位 C_i 要经过 $3.5t_y$ 延时。

2. 4 位先行进位电路 CLA

提高加法器运算速度的关键是缩短串行进位中进位逐位的传递时间, 让各位进位同时产生。在分组先行进位方式中, 组内采用并行方式, 假设 4 位为一组, 4 个进位输出信号仅由进位产生函数 G_i 、进位传递函数 P_i 以及最低位进位 C_0 决定, 所以这些进位信号是同时产生的。

$$C_1 = G_1 + P_1 C_0$$

$$C_2 = G_2 + P_2 C_1 = G_2 + P_2 G_1 + P_2 P_1 C_0$$

$$C_3 = G_3 + P_3 C_2 = G_3 + P_3 G_2 + P_3 P_2 G_1 + P_3 P_2 P_1 C_0$$

$$C_4 = G_4 + P_4 C_3 = G_4 + P_4 G_3 + P_4 P_3 G_2 + P_4 P_3 P_2 G_1 + P_4 P_3 P_2 P_1 C_0$$

实现上述进位逻辑函数的电路称之为 4 位先行进位电路 (carry look ahead, CLA), 如图 4-3 所示。

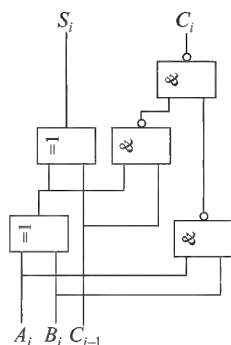


图 4-2 全加器的逻辑图

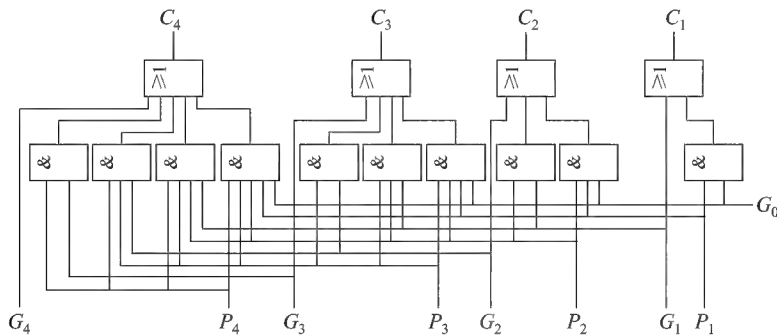


图 4-3 4 位先行进位电路

若用 4 位 CLA 电路组成 CLA 加法器, 还必须配上进位产生/传递电路和求和电路。进位产生/传递电路是为了产生 G_i 和 P_i , 而求和电路则是用来产生各位的和 (S_i)。因为进位产生/传递电路的延迟时间是 $1.5t_y$, 求和电路的延迟时间是 $3t_y$ 。所以用以上 3 种电路组成 16 位字长的单级 CLA 加法器, 其总的延迟时间 (包括 P_i 、 G_i 产生时间、进位延迟时间和求和时间) 为:

$$T = 1.5t_y + 4 \times 2t_y + 3t_y = 12.5t_y$$

其中, $4 \times 2t_y$ 为 4 个 4 位 CLA 电路的进位传递时间。进一步分析可以看到, 由 $C_0 \rightarrow C_4$ 的最大延迟时间为 $1.5t_y + 2t_y = 3.5t_y$, 由 $C_0 \rightarrow C_{16}$ 的最长延迟时间为 $(1.5 + 4 \times 2)t_y = 9.5t_y$ 。

3.4 位成组先行进位电路

为了要产生组进位函数,需要对原来的 CLA 电路进行修改:

第1小组内产生 G_1^* 、 P_1^* 、 C_3 、 C_2 、 C_1 ,不产生 C_4 ;

第2小组内产生 G_2^* 、 P_2^* 、 C_7 、 C_6 、 C_5 ,不产生 C_8 ;

第3小组内产生 G_3^* 、 P_3^* 、 C_{11} 、 C_{10} 、 C_9 ,不产生 C_{12} ;

第4小组内产生 G_4^* 、 P_4^* 、 C_{15} 、 C_{14} 、 C_{13} ,不产生 C_{16} 。

这种电路称为成组先行进位部件(block carry look ahead,BCLA),图4-4为第1组的BCLA电路。

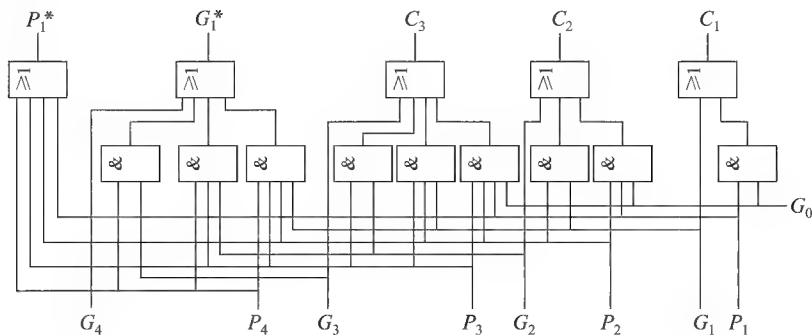


图4-4 4位BCLA电路

利用4位BCLA电路再配上前述的进位产生/传递电路与求和电路可组成BCLA加法器。整个加法器的总延迟时间为:

$$T=1.5t_y+6t_y+3t_y=10.5t_y$$

4.4 4位CLA加法器和4位BCLA加法器的比较

4位CLA加法器是由4位加法器配上CLA电路组成的,而4位BCLA加法器是由4位加法器配上BCLA电路组成的。两者都能实现4位操作数的加法运算,其区别在于,前者除产生4位和 $S_4 \sim S_1$ 外,还将产生4位加法器向高位的进位 C_4 ;后者除产生4位和 $S_4 \sim S_1$ 外,不产生向高位的进位 C_4 ,但产生组进位产生函数 G_i^* 和组进位传递函数 P_i^* 。如图4-5所示。

5. 原码和反码的加减运算

前面已经提到,在通用计算机中,通常采用补码实现加、减、乘、除运算。原码和反码的加减运算要比补码的加减运算复杂不少,下面讨论原码和反码的加减运算。

(1) 原码加减运算

对原码表示的两个数进行加减运算时,计算机的实际操作是加还是减,不仅取决于指令的操作码,还取决于两个操作数的符号,例如:加法时可能要做减法(两数异号);减法时

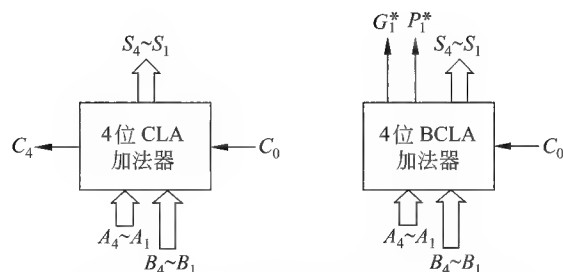


图 4 5 4 位 CLA 加法器和 4 位 BCLA 加法器

又可能做加法(两数异号),所以原码加减运算的实现是比较复杂的。

设有两个定点数:

$$\text{被加(减)数}[X]_{\text{原}} = X_s, X_1 X_2 \cdots X_n$$

$$\text{加(减)数}[Y]_{\text{原}} = Y_s, Y_1 Y_2 \cdots Y_n$$

$$\text{则两数之和(差)为:}[S]_{\text{原}} = S_s, S_1 S_2 \cdots S_n$$

两个操作数的加减运算有 8 种可能的组合,它们可以组合归并为 4 类实际操作:

正数 + 正数 = 正数 正数 - 负数

负数 + 负数 = 负数 负数 - 正数

正数 + 负数 = 正数 正数 - 正数

负数 + 正数 = 负数 负数 - 负数

前两类是同号相加和异号相减,实际操作是做绝对值相加,结果符号取被加(减)数的符号。后两类是异号相加和同号相减,实际操作为绝对值相减,结果符号与绝对值大的数的符号相同。

指令加、减和机器实际加、减是两个不同的概念,机器加、减与指令加、减的关系式如下:

$$[\text{机器加}] = (X_s \oplus Y_s)[\text{指令加}] + (\overline{X_s \oplus Y_s})[\text{指令减}]$$

$$[\text{机器减}] = (X_s \oplus Y_s)[\text{指令加}] + (\overline{X_s \oplus Y_s})[\text{指令减}]$$

加减运算结果的符号表达式为:

$$\begin{aligned} S_s &= [\text{机器减}](|X| < |Y|) \overline{X_s} + [\text{机器减}](|X| < |Y|) X_s \\ &= ([\text{机器减}] (|X| < |Y|)) \oplus X_s \end{aligned}$$

在大多数计算机中,通常只设置加法器而不设置减法器,因此减法运算将转换为加法运算来实现。原码运算时,用 $|X| + [|Y|]_{\text{变补}}$ 来代替 $|X| - |Y|$ 。

原码加减运算规则如下:

① 参加运算的操作数取其绝对值。

② 若做加法,则两数直接相加,若做减法,则将减数先变一次补,再进行加法运算。

- ③ 运算之后,可能有两种情况:
- a. 有进位,结果为正,即得到正确的结果;
 - b. 无进位,结果为负,则应再变一次补,才能得到正确的结果。
- ④ 加上符号位,得到用原码表示的结果。
- 通常,把运算之前的变补称为前变补,运算之后的变补称为后变补。

例 4-2 $12-9=3$ 。

$$\begin{array}{rcl}
 1100 & \text{——} & 12 \\
 + 0111 & \text{——} & \text{对 9 变补(前变补)} \\
 \hline
 0011 & \text{——} & \text{结果为 3,有进位,表示结果为正} \\
 0\ 0011 & \text{——} & \text{加符号}
 \end{array}$$

例 4-3 $9-12=-3$ 。

$$\begin{array}{rcl}
 1001 & \text{——} & 9 \\
 + 0100 & \text{——} & \text{对 12 变补(前变补)} \\
 \hline
 1101 & \text{——} & \text{无进位,表示结果为负} \\
 0011 & \text{——} & \text{后变补,结果为 3} \\
 1\ 0011 & \text{——} & \text{加符号}
 \end{array}$$

(2) 反码加减运算

与补码加减运算类似,反码加减运算应有:

$$\begin{aligned}
 [X+Y]_{\text{反}} &= [X]_{\text{反}} + [Y]_{\text{反}} \\
 [X-Y]_{\text{反}} &= [X]_{\text{反}} + [-Y]_{\text{反}}
 \end{aligned}$$

反码加减运算规则如下:

- ① 参加运算的操作数用反码表示;
- ② 符号位作为数的一部分参加运算;
- ③ 若做加法,则两数直接相加,若做减法,则将被减数与连同符号位一起变反后的减数相加;
- ④ 运算时如果符号位产生进位,则在末位加“1”,称为循环进位;
- ⑤ 结果以反码表示。

例 4-4 $A=0.1001, B=-0.0100$,求 $[A+B]_{\text{反}}$ 。

因为 $[A]_{\text{反}}=0.1001, [B]_{\text{反}}=1.1011$

$$\begin{array}{r}
 0.1001 \\
 + 1.1011 \\
 \hline
 10.0100 \\
 \begin{array}{c} \text{└───┐} \\ \text{└───┘} \end{array} \rightarrow 1 \\
 \hline
 0.0101
 \end{array}$$

所以 $[A+B]_{\text{反}}=0.0101, A+B=0.0101$

例 4-5 $A=0.1001, B=-0.0100$, 求 $[A-B]_{\text{反}}$ 。

因为 $[A]_{\text{反}}=0.1001, [B]_{\text{反}}=1.1011, [-B]_{\text{反}}=0.0100$

$$\begin{array}{r} 0.1001 \\ + 0.0100 \\ \hline 0.1101 \end{array}$$

所以 $[A-B]_{\text{反}}=0.1101, A-B=0.1101$

6. 补码的移位操作

算术移位时应保持数的符号位不变,而数值的大小则要发生变化。左移 1 位相当于乘以 2,右移 1 位相当于除以 2。

例 4-6 设 $[X]_{\text{补}}=X_s.X_1X_2\cdots X_n$, 求证: $\left[\frac{1}{2}X\right]_{\text{补}}=X_s.X_sX_1X_2\cdots X_n$,

证明: 因为 $X=-X_s+\sum_{i=1}^n X_i2^{-i}$

$$\begin{aligned} \text{所以 } \frac{1}{2}X &= -\frac{1}{2}X_s + \frac{1}{2}\sum_{i=1}^n X_i2^{-i} = -X_s + \frac{1}{2}X_s + \frac{1}{2}\sum_{i=1}^n X_i2^{-i} = -X_s + \\ &\frac{1}{2}\sum_{i=0}^n X_i2^{-(i+1)} \end{aligned}$$

根据补码与真值的关系有:

$$\left[\frac{1}{2}X\right]_{\text{补}}=X_s.X_sX_1X_2\cdots X_n$$

7. 各种舍入方法的比较

减少运算中精度损失的关键是要处理好运算中尾数超出字长的部分,使之精度损失小。

为了对不同的舍入处理方法做对比,使用误差曲线,并以尾数基数 $r=2$,尾数位数 $m=2$ 为例来讨论。图 4-6 中横坐标是处理前的实际值,纵坐标是经舍入处理后的结果值,虚线为理想的无精度损失曲线。

(1) 恒舍法

其方法是将尾数超出机器字长的部分截去,误差曲线如图 4-6(a)所示。对于正数总是产生负误差,除非那些圆点处才无误差。

这种方法的好处是实现最简单,不增加硬件,不需要处理时间,但由于最大误差较大,平均误差大且无法调节,因而已很少使用。

(2) 恒置 1 法

其方法是令机器运算的规定字长的最低位恒为 1,误差曲线如图 4-6(b)所示。对于正数,误差有正有负(如 $11|10\cdots1$ 舍入成 $11|$,造成负误差; $10|10\cdots1$ 舍入成 $11|$,造成正

误差;11|00...0 舍入成 11|,无误差)。统计平均误差接近于零但略偏正,平均误差无法调节。

这种方法的好处是实现简单,不需要增加硬件和处理时间,平均误差趋于零。主要缺点是最大误差最大,比恒舍法还大。

(3) 舍入法

舍入法又称下舍上入法(0 舍 1 入法),误差曲线如图 4-6(c)所示。对于正数,误差有正有负(如 10|01...1 舍入成 10|,造成负误差;10|10...0 舍入成 11|,造成正误差;01|00...0 舍入成 01|,无误差)。统计平均误差趋于零但略偏正,平均误差无法调节。

这种方法的好处是实现简单,增加的硬件很少,最大误差小,平均误差接近于零。主要缺点是处理速度慢,最坏的情况下可能需要从尾数最低位进位至最高位。

(4) 查表舍入法

查表舍入法的误差曲线如图 4-6(d)所示。这种方法速度较快,平均误差可调节到零,但缺点是需要的硬件量大,不过随着器件价格的下降和集成度的改进,其使用将会增多。

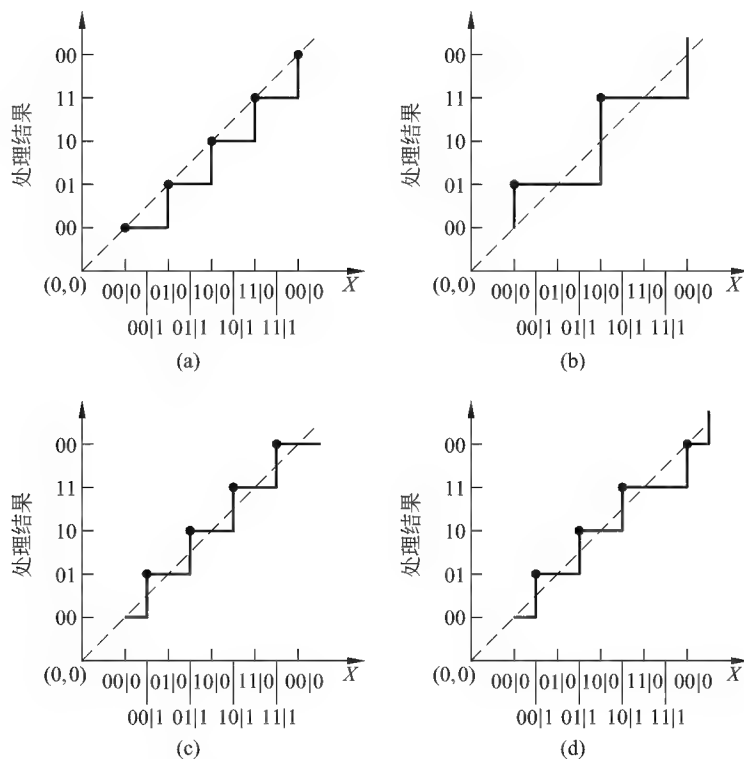


图 4-6 各种舍入处理方法的误差曲线

8. 补码校正乘法

补码乘法不能简单的套用原码乘法的算法,这是因为补码的符号位是参加运算的。所谓校正法是将 $[X]_{\text{补}}$ 和 $[Y]_{\text{补}}$ 按原码规则运算,所得结果根据情况再加以校正,从而得到 $[X \times Y]_{\text{补}}$ 。下面分两种情况讨论。

(1) 被乘数 X 的符号任意,乘数 Y 为正数

因为 $[X]_{\text{补}} = X_s . X_1 X_2 \cdots X_n = 2 + X \pmod{2}$

$$[Y]_{\text{补}} = 0 . Y_1 Y_2 \cdots Y_n = Y$$

所以 $[X]_{\text{补}} \times [Y]_{\text{补}} = [X]_{\text{补}} \times Y = (2 + X) \times Y = 2Y + X \times Y$

由于 Y 是大于 0 的正数,根据模运算的性质,有 $2Y = 2 \pmod{2}$ 。

所以 $[X]_{\text{补}} \times [Y]_{\text{补}} = 2 + XY = [X \times Y]_{\text{补}} \pmod{2}$

可见当乘数 $Y > 0$ 时,不管被乘数 X 的符号如何都可直接按原码乘法运算,只是移位时按补码规则进行。

(2) 被乘数 X 的符号任意,乘数 Y 为负数

因为 $[X]_{\text{补}} = X_s . X_1 X_2 \cdots X_n = 2 + X \pmod{2}$

$$[Y]_{\text{补}} = 1 . Y_1 Y_2 \cdots Y_n = 2 + Y \pmod{2}$$

$$Y = [Y]_{\text{补}} - 2 = 1 . Y_1 Y_2 \cdots Y_n - 2 = 0 . Y_1 Y_2 \cdots Y_n - 1$$

所以 $XY = X \times (0 . Y_1 Y_2 \cdots Y_n) - X$

$$[X \times Y]_{\text{补}} = [X \times (0 . Y_1 Y_2 \cdots Y_n)]_{\text{补}} + [-X]_{\text{补}}$$

因为 $(0 . Y_1 Y_2 \cdots Y_n) > 0$

所以 $[X \times Y]_{\text{补}} = [X]_{\text{补}} \times (0 . Y_1 Y_2 \cdots Y_n) + [-X]_{\text{补}}$

可见当乘数 $Y < 0$ 时,可以先把 $[Y]_{\text{补}}$ 的符号位丢掉不管,仍按原码乘法运算,最后再加上 $[-X]_{\text{补}}$ 进行校正。

将上述两种情况综合起来,就得到了补码乘法的统一表达式:

所以 $[X \times Y]_{\text{补}} = [X]_{\text{补}} \times (0 . Y_1 Y_2 \cdots Y_n) + [-X]_{\text{补}} \times Y_s$

9. 原码两位乘法

为了提高乘法的执行速度,可以选用两位乘法的方案。在主教材中已经讨论了补码两位乘法,在这里介绍原码两位乘法。

原码两位乘法和原码一位乘法一样,符号位单独处理。乘数的相邻两位 $Y_{i-1}Y_i$ 有 4 种状态,决定进行何种操作。

$Y_{i-1}Y_i = 00$, 相当于 $0 \times X$, 部分积 $+0$, 右移 2 位。

$Y_{i-1}Y_i = 01$, 相当于 $1 \times X$, 部分积 $+X$, 右移 2 位。

$Y_{i-1}Y_i = 10$, 相当于 $2 \times X$, 部分积 $+2X$, 右移 2 位。

$Y_{i-1}Y_i = 11$, 相当于 $3 \times X$, 部分积 $+3X$, 右移 2 位。

其中, $+3X$ 的运算,用普通的加法器不能一次完成,如果分为两次执行,则又降低了速

度。可将 $3X$ 当作 $(4X-X)$ 来处理,本次操作执行 $-X$,用一个欠账触发器 C_j 记下欠账,下一次操作时再补上 $+4X$ 。由于本次累加后部分积要右移 2 位,从相对关系来看,相当于被乘数左移了 2 位,因而下一次实际上只需执行 $+X$,就等于前次完成了 $+4X$ 操作,下面通过一个实例说明这一关系。

设: $A=0.0001,X=0.0101$ 。

$2^{-2}(A+4X)$ 为:

$$\begin{array}{r} 000.0001 \qquad A \\ + \quad 001.0100 \qquad 4X \\ \hline 001.0101 \\ 2 \rightarrow 000.010101 \end{array}$$

$2^{-2}A+X$ 为:

$$\begin{array}{r} 000.0001 \qquad A \\ 2 \rightarrow 000.000001 \\ + \quad 000.0101 \qquad X \\ \hline 000.010101 \end{array}$$

所以本次 $2^{-2}(A+4X)=$ 下次 $(A'+X)$

其中, A' 为已右移了 2 位的 A ,即 $A'=2^{-2}A$ 。

原码两位乘法规则:

- ① 参加运算的操作数取其绝对值;
- ② 符号位单独处理 $P_s=X_s\oplus Y_s$;
- ③ 欠账触发器 C_j 初始值为 0;
- ④ 根据乘数的最低两位 $Y_{n-1}Y_n$ 和欠账触发器 C_j 的值决定每次应执行的操作,见表 4-1;
- ⑤ $-|X|$ 通过 $+[-X]_{\text{变补}}$ 实现,所以右移按补码规则进行;
- ⑥ 当乘数的数值位为 n 位(不连符号位),应作 $n/2$ 次累加和移位,如有欠账,再做一次加法。

表 4-1 原码两位乘法运算操作

$Y_{n-1}Y_n$	C_j	操 作
00	0	部分积+0,右移 2 位, $0 \rightarrow C_j$
00	1	部分积+ $ X $,右移 2 位, $0 \rightarrow C_j$
01	0	部分积+ $ X $,右移 2 位, $0 \rightarrow C_j$
01	1	部分积+ $2 X $,右移 2 位, $0 \rightarrow C_j$

续表

$Y_{n-1}Y_n$	C_j	操 作
10	0	部分积 $+2 X $, 右移 2 位, $0 \rightarrow C_j$
10	1	部分积 $- X $, 右移 2 位, $1 \rightarrow C_j$
11	0	部分积 $- X $, 右移 2 位, $1 \rightarrow C_j$
11	1	部分积 $+0$, 右移 2 位, $1 \rightarrow C_j$

由于在运算中有 $+2|X|$, 累加时产生的进位可能侵占符号位, 所以被乘数和部分积应取 3 个符号位。乘数需凑足偶数位, 以便于两位一组的运算, 由于最后可能会有欠账, 故乘数应取双符号位, 以便最后一次能处理前面留下的欠账(出现“001”代码)。实际上乘数不取符号位也可以, 但要记住还清欠账。

注意不要将原码两位乘法 and Booth 乘法相混淆, C_j 是欠账触发器, 它是由前次操作是否有欠账来决定置位或复位的, 而不像 Booth 乘法中的 Y_{n+1} 是由乘数直接右移得到的。另外每次得到的部分积也不同, 前者表示每次得到两位乘数的部分积, 后者只得到一位乘数的部分积。

例 4-7 已知: $X = -0.111111$, $Y = 0.111001$, 利用原码两位乘法求: $X \times Y$ 。

解: $|X| = 000.111111 \rightarrow B$, $|Y| = 000.111001 \rightarrow C$, $0 \rightarrow A$

$[|X|]_{\text{变补}} = 111.000001$, $2|X| = 001.111110$

A	C	C_j	说明
000.000000	00.111001	0	
$+ X $ 000.111111			$C_6C_7C_j=010$, $+ X $, $0 \rightarrow C_j$
000.111111			
$2 \rightarrow$ 000.001111	11	0011100	部分积右移 2 位
$+2 X $ 001.111110			$C_6C_7C_j=100$, $+2 X $, $0 \rightarrow C_j$
010.001101			
$2 \rightarrow$ 000.100011	0111	001110	部分积右移 2 位
$- X $ 111.000001			$C_6C_7C_j=110$, $- X $, $1 \rightarrow C_j$
111.100100			
$2 \rightarrow$ 111.111001	000111	001	部分积右移 2 位
$+ X $ 000.111111			$C_6C_7C_j=001$, $+ X $, 还欠账
000.111000			

$$P_s = X_s \oplus Y_s = 1 \oplus 0 = 1$$

所以 $[X \times Y]_{\text{原}} = 1.111000000111$
 $X \times Y = -0.11100000011$

10. 不同情况除法运算中的寄存器安排

在主教材中,我们已经讨论了原码、补码的恢复余数法和不恢复余数除法。在这几种除法算法中,被除数和除数都是定点小数,且数值位的位数都为 n 。如果被除数为 $2n$ 位(双倍字长)或被除数和除数都是定点整数,前述的算法还适用吗? 应当说除法的算法是适用的,但是在除法运算中要用到的 3 个寄存器的安排上有些变化。不同情况下寄存器的安排如表 4-2 所示。

表 4-2 不恢复余数除法运算时寄存器的安排

操作数类型		A 寄存器		B 寄存器	C 寄存器	
		初态	终态		初态	终态
定点 小数	单字长	被除数	→(部分余数)→余数	除数	0	→商
	双字长	被除数高位	→(部分余数)→余数	除数	被除数低位	→商
定点 整数	单字长	0	→(部分余数)→余数	除数	被除数	→商
	双字长	被除数高位	→(部分余数)→余数	除数	被除数低位	→商

若用双字长 $2n$ 位被除数除以 n 位的除数,得到 n 位的商数,这种除法通常称为双精度除法。当被除数为双字长时,被除数高位部分存放在寄存器 A 中,低位部分存放在寄存器 C 中,其余同前述的单精度除法。

在进行整数除法时,必须满足 $|被除数| \geq |除数|$ 的条件,同时寄存器的分配也与进行小数除法时有所不同。若参加运算的操作数是整数,则在运算初始时,寄存器 A 的初值为 0,寄存器 B 用于存放除数,寄存器 C 用于存放被除数。除法结束时,A 中存放余数,C 中存放商,B 中内容不变。

11. 二进制移码加减法

在浮点数据表示时,阶码通常使用移码来表示。我们已经知道,两个 $n+1$ 位、偏置值为 2^n 的移码在做加减运算时,操作数用移码表示,结果也用移码表示。直接利用移码运算后的结果需要进行必要的修正,即有:

$$[A+B]_{\text{移}} = [A]_{\text{移}} + [B]_{\text{移}} - 2^n$$

$$[A-B]_{\text{移}} = [A]_{\text{移}} + [-B]_{\text{移}} + 2^n$$

由于此时移码与补码的不同仅在于两者的最高位(符号位)不同,故有:

$$[A]_{\text{移}} = [A]_{\text{补}} + 2^n$$

所以在进行移码加减运算时,通常使用如下公式:

$$[A+B]_{\text{移}} = [A]_{\text{移}} + [B]_{\text{补}}$$

$$[A-B]_{\text{移}} = [A]_{\text{移}} + [-B]_{\text{补}}$$

为了便于判断溢出,移码采用2位符号位(变形移码):第一位符号为0,而第二位代表数据的正负。即当A为正数时, $[A]_{\text{移}}$ 的符号为01;而当A为负数时, $[A]_{\text{移}}$ 的符号为00。变形移码只是在运算过程中采用,在传送和存储时仍只保留1位符号位。

因此,移码加减运算规则可归纳如下:

- ① 参加运算的两个操作数均用移码表示;
- ② 采用2位符号位,即用变形移码表示;
- ③ 符号位作为数的一部分参加运算;
- ④ 运算结果以移码表示,若第一位符号为0,结果正常;若第一位符号为1,表示溢出;符号位为10时表示正溢出,符号位为11时表示负溢出。

例 4-8 $A=1011, B=-1110$, 求: $[A+B]_{\text{移}}$ 。

解: 因为 $[A]_{\text{移}}=011011$ $[B]_{\text{补}}=110010$

$$\begin{array}{r} 011011 \\ + 110010 \\ \hline 001101 \end{array} \quad \begin{array}{l} [A]_{\text{移}} \\ [B]_{\text{补}} \\ [A+B]_{\text{移}} \end{array}$$

所以 $[A+B]_{\text{移}}=001101, A+B=-0011$

例 4-9 $A=1011, B=-0010$, 求: $[A-B]_{\text{移}}$ 。

解: 因为 $[A]_{\text{移}}=011011$ $[B]_{\text{补}}=111110$ $[-B]_{\text{补}}=000010$

$$\begin{array}{r} 011011 \\ + 000010 \\ \hline 011101 \end{array} \quad \begin{array}{l} [A]_{\text{移}} \\ [-B]_{\text{补}} \\ [A-B]_{\text{移}} \end{array}$$

所以 $[A-B]_{\text{移}}=011101, A-B=1101$

例 4-10 $A=-1101, B=-1010$, 求: $[A+B]_{\text{移}}$ 。

解: 因为 $[A]_{\text{移}}=000011$ $[B]_{\text{补}}=110110$

$$\begin{array}{r} 000011 \\ + 110110 \\ \hline 111001 \end{array} \quad \begin{array}{l} [A]_{\text{移}} \\ [B]_{\text{补}} \\ [A+B]_{\text{移}} \end{array}$$

结果为负溢出。

例 4-11 $A=1101, B=-1010$, 求: $[A-B]_{\text{移}}$ 。

解: 因为 $[A]_{\text{移}}=011101$ $[B]_{\text{补}}=110110$ $[-B]_{\text{补}}=001010$

$$\begin{array}{r} 011101 \\ + 001010 \\ \hline 100111 \end{array} \quad \begin{array}{l} [A]_{\text{移}} \\ [-B]_{\text{补}} \\ [A-B]_{\text{移}} \end{array}$$

结果为正溢出。

12. 浮点乘法运算的溢出和舍入问题

浮点乘法运算需要做阶码相加。同号相加,若为正阶码,则可能上溢;若为负阶码,则可能下溢。如何正确地判断出上溢和下溢呢?下面首先分别讨论下溢和上溢的问题。

(1) 判断下溢

产生乘法下溢有两种可能:一是求乘积的阶码时已下溢,二是乘积左规时阶码减1而造成下溢。这样,就有一个什么时候判下溢的问题。

例 4-12 已知: $A = -1 \times 2^{128}$, $B = -1 \times 2^{-1}$, 求: $A \times B$ 。

解: 假设 A 、 B 的阶码和尾数均采用补码表示,阶码取 9 位(包括两位符号位)。则有:

$$[A]_{\text{补}} = 110000000, 11.00 \cdots 0; [B]_{\text{补}} = 111111111, 11.00 \cdots 0$$

按照运算规则:

$$[A \times B]_{\text{补}} = 101111111, 01.00 \cdots 0$$

此时,两数阶码之和为 101111111(即-129),尾数之积为 01.00...0(即+1)。如果在阶码求和之后就判溢出,则此时被判为下溢。但实际上计算结果需要右规,阶码+1,最后阶码为 110000000(即-128),尾数为 00.10...0,结果没有溢出。

例 4-12 说明,如果在阶码求和后就判溢出,可能出现本没有溢出,而被误认为是下溢的情况,错误地扩大了溢出范围。因此正确的做法是在规格化后判断溢出,这才不会扩大溢出范围,但这种做法也有可能造成判断错误。

例 4-13 已知: $A = 0.5 \times 2^{128}$, $B = 0.5 \times 2^{128}$, 求: $A \times B$ 。

解: $[A]_{\text{补}} = 110000000, 00.10 \cdots 0; [B]_{\text{补}} = 110000000, 00.10 \cdots 0$

按照运算规则:

$$[A \times B]_{\text{补}} = 100000000, 00.01 \cdots 0$$

此时,两数阶码之和为 100000000(即-256),尾数之积为 00.01...0,此时阶符为 10,应该判断为下溢。计算结果左规后,积的尾数变成 00.10...0。阶码需要-1(+111111111):

$$100000000 + 111111111 = 011111111$$

规格化后,由于阶码减1,阶符为 01,变成了上溢的形式。

例 4-13 说明,如果在规格化后判断溢出,就可能把本来是下溢错判为上溢了。

如何解决这个矛盾呢?一个简单的方法是:用求阶码和后的阶码寄存器的最高位(代表阶符)来参与控制。设求阶码和后,阶码寄存器的内容为:

$$R_{E8} \ R_{E9} \ R_{E1} \ R_{E2} \ R_{E3} \ R_{E4} \ R_{E5} \ R_{E6} \ R_{E7}$$

左规时,阶码加法器的输出为:

$$F_{E8} \ F_{E9} \ F_{E1} \ F_{E2} \ F_{E3} \ F_{E4} \ F_{E5} \ F_{E6} \ F_{E7}$$

当 $R_{ES}=1$, 且 $F_{ES} \neq F_{E0}$ 时下溢, 即下溢条件 $= R_{ES} \cdot (F_{ES} \oplus F_{E0})$

这样, 在规格化后判下溢, 既不会扩大溢出范围, 也不会错判成上溢了。

判断出下溢后, 应相应地将下溢标志触发器置 1, 并将下溢中断标志位置 1, 以便在规格化之后, 将乘法结果清为机器零。

(2) 判断上溢

产生乘法上溢也有两种可能: 一是求乘积的阶码时已上溢, 二是乘积右规时阶码加 1 而造成上溢。这样, 也有一个什么时候判上溢的问题。

例 4-14 已知: $A=0.5 \times 2^{127}$, $B=0.5 \times 2^1$, 求: $A \times B$ 。

解: $[A]_{\text{补}}=001111111, 00.10 \cdots 0$; $[B]_{\text{补}}=000000001, 00.10 \cdots 0$

按照运算规则:

$$[A \times B]_{\text{补}}=010000000, 00.01 \cdots 0$$

此时, 两数阶码之和为 010000000 (即 128), 尾数之积为 00.01...0。如果在阶码求和之后就判断溢出, 则此时被判断为上溢。但实际上, 计算结果需要左规, 最后阶码为 001111111 (即 127), 尾数为 00.10...0, 结果没有溢出。

例 4-14 说明, 如果在阶码求和后就判断溢出, 可能出现本不应为溢出, 而被误认为是上溢, 错误地扩大了溢出范围。因此正确的做法是应该在规格化后判溢出。此时并不存在类似下溢时出现的问题, 不会把上溢错判断为下溢。

例 4-15 已知: $A=-1 \times 2^{127}$, $B=-1 \times 2^{127}$, 求: $A \times B$ 。

解: $[A]_{\text{补}}=001111111, 11.00 \cdots 0$; $[B]_{\text{补}}=001111111, 11.00 \cdots 0$

按照运算规则:

$$[A \times B]_{\text{补}}=011111110, 01.00 \cdots 0$$

此时, 两数阶码之和为 011111110 (即 254), 尾数之积为 01.00...0, 此时阶符为 01, 应该判断为上溢。计算结果右规后, 积的尾数变成 00.10...0。阶码需要 +1, 修正为 011111111 (即 255), 仍保持上溢的形式, 不会错判为下溢。

例 4-15 说明, 如果在规格化之后判溢出, 并不会出现误判。

但是下溢时已用 $R_{ES} \cdot (F_{ES} \oplus F_{E0})$ 来判断, 如果上溢时只用 $F_{ES} F_{E0}$ 来判断, 则类似例 4-13 中出现的下溢就可能误判为上溢。因此只有用类似于判下溢的方法来判上溢, 才不会与下溢相混淆, 也就是不会出错。所以上溢的条件 $= R_{ES} \cdot (F_{ES} \oplus F_{E0})$

实际上, 只会出现 $R_{ES} F_{ES} F_{E0}$ 的情况, 而不会出现 $R_{ES} \overline{F_{ES}} F_{E0}$ 的情况, 只是为使判断上溢与判断下溢统一而已。

同样, 判断出上溢后, 也应相应地将上溢标志触发器置 1, 并将上溢中断标志位置 1, 以便进行上溢中断的处理。

由于参加运算的数为规格化的数, 因而乘积的尾数的绝对值必定大于等于 1/4, 所以即使需要左规, 最多只需一次, 且左规无舍入问题。由于补码 $[-1]_{\text{补}}$ 是有意义的, 所以,

当两数尾数都为 $[-1]_{补}$ 时,尾数相乘后为 $01.00\cdots 0$ (即 $+1$),此时需要右规,右规也只能一次。由于右移1位并未丢掉尾数,所以也不需舍入。乘法的舍入只发生在对乘积不取双倍字长的尾数,而取单字长时,为确保一定乘积精度时的舍入处理。

13. 多功能算术逻辑单元 74181

74181 是对前述的 4 位先行进位加法器进行修改而得到的。74181 可以实现多种算术运算和逻辑运算,由功能选择线 $S_3 \sim S_0$ 和操作方式 M 控制。为了和选择信号 S_i 有所区别,将原先行进位加法器的输出和 S_i 改为 F_i ,并且进位 C_i 受 M 的控制。经过修改后,和及进位的公式变成:

$$F_i = A_i \oplus B_i \oplus C_{i-1} = P_i \oplus C_{i-1}$$

$$C_i = G_i \cdot M + P_i \cdot M \cdot C_{i-1}$$

74181 还提供了 3 个信号, G 、 P 和 $\overline{C_{n+4}}$,供级联更多位数的 ALU 使用。 G 和 P 即 4 位先行进位加法器的组进位产生函数和组进位传递函数, $\overline{C_{n+4}}$ 即 $\overline{C_n}$ 。74181 的逻辑图如图 4-7 所示。

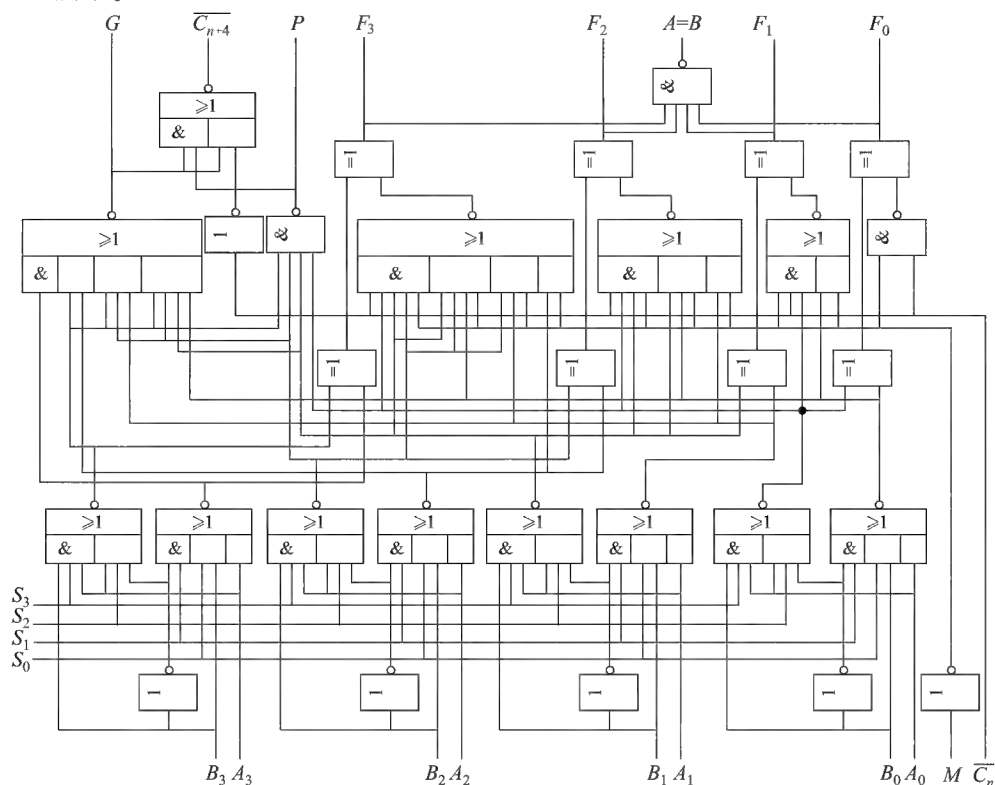


图 4-7 74181 逻辑图

74181 的结构适合于将它们级联成各种位数的 ALU。每片 74181 可作为一个 4 位先行进位加法器。当加法器采用组间串行进位时,利用 C_{n+4} 输出端,可将多个 74181 串联,组成字长是 4 的倍数的 ALU。当加法器采用两级先行进位时,利用 G 、 P 输出端,此时另需一片 74182 —— 先行进位发生器。

74181 除能实现 16 种算术、逻辑运算功能外,还能实现很多比较功能,例如: $=$ 、 $>$ 、 \geq 、 $<$ 、 \leq 、 \neq 。这些功能见表 4-3。检查 $A=B$ 输出端和进位输出 C_{n+4} 的值,选择一定的操作,就可以决定 A 和 B 的相对大小。

表 4 3 用 74181 执行比较操作

输出	状态	操作	负逻辑	正逻辑	备注
$A=B$	1	A 减 B	$A=B$	$A=(B$ 减 $1)$	
	1	$A \oplus B$	$A \neq B$	$A=B$	
	1	$A \oplus B$	$A=B$	$A \neq B$	
C_{n+4}	1	A 减 B	$A \geq B$	$A < B$	有进位输入时
	0	A 减 B	$A < B$	$A \geq B$	
	1	A 减 B 减 1	$A > B$	$A \leq B$	无进位输入时
	0	A 减 B 减 1	$A \leq B$	$A > B$	

14. 先行进位发生器 74182

图 4-8 是 74182 的逻辑图,图中 C_{n+x} 、 C_{n+y} 、 C_{n+z} 是 3 个进位输出信号, G 、 P 是大组进位产生函数和进位传递函数。

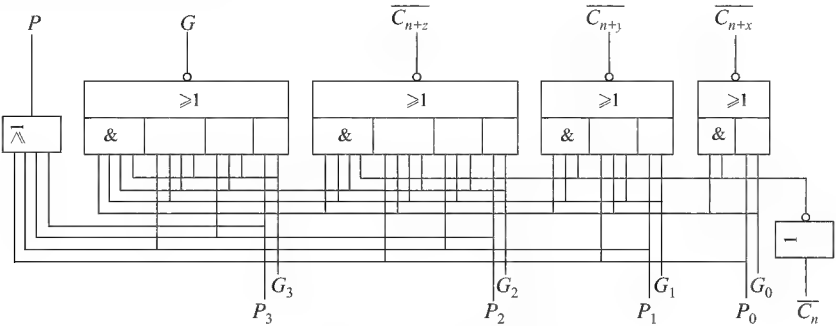


图 4-8 74182 逻辑图

$$\begin{aligned} C_{n+x} &= G_0 (P_0 + C_n) \\ C_{n+y} &= G_1 (P_1 + G_0 (P_0 + C_n)) \\ C_{n+z} &= G_2 (P_2 + G_1 (P_1 + G_0 (P_0 + C_n))) \end{aligned}$$

$$G = G_3(P_3 + G_2)(P_3 + P_2 + G_1)(P_3 + P_2 + P_1 + G_0)$$

$$P = P_3 + P_2 + P_1 + P_0$$

图 4-8 中的 $G_0 \sim G_3$ 即 $G_0^* \sim G_3^*$, $P_0 \sim P_3$ 即 $P_0^* \sim P_3^*$, G 即 G^{**} , P 即 P^{**} 。

74181 的 4 位作为一个小组,小组间既可以采用串行进位,也可以采用并行进位。当采用串行进位时,只要把低一片的 C_{n+4} 与高一片的 C_n 相连即可。当采用组间并行进位时,需要增加一片 74182,74182 的输出 C_{n-x} 、 C_{n+y} 、 C_{n+z} 分别接前 3 片 74181 的 C_n 端。

15. 位片式运算器

采用大规模集成电路技术可将 n 位寄存器组、 n 位选择器、 n 位 ALU、 n 位移位器等集成在一块芯片上,成为一片 n 位运算器。将若干块这样的位片连接起来,就能构成较长位数的运算器。这种方法使系统组成灵活方便,且可大批量生产位片。代表性的位片有 AMD2900/29000/29300 序列。图 4-9 是 AMD2900 系列位片的粗框图。

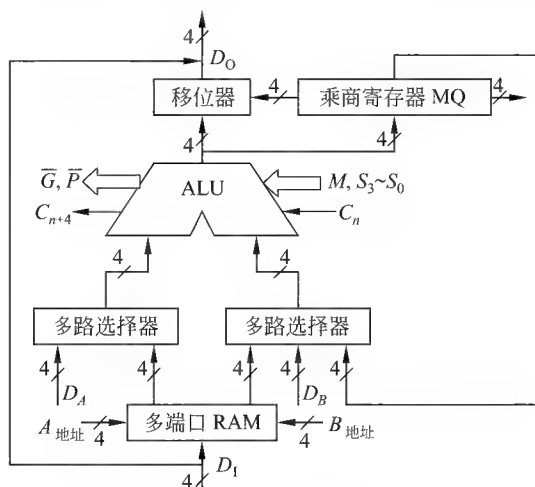


图 4-9 位片式运算器组成

双端口随机存储器(RAM)构成一个 16×4 位的通用寄存器组。所谓双端口,是指可以同时向它送入两个地址: A 地址和 B 地址,因而可同时选中两个寄存器,它们同时将各自的 4 位数据送往多路选择器,供 ALU 运算处理。

ALU 类似于 74181 的逻辑结构,在此基础上进一步扩展了功能,可实现乘、除运算。它的功能控制信号有 M 、 S_3 S_2 S_1 S_0 、..., 进位输入 C_n , 进位输出 C_{n+4} , 进位辅助函数 P 、 G 等,此外还输出某些状态信息。

乘商寄存器 MQ 用于乘、除运算。在乘法运算时用来存放乘数,运算结束时存放乘积的低位部分;在除法运算时用来存放商。 MQ 寄存器也可作为辅助寄存器使用。

多路选择器实现 ALU 的输入选择。它的信息来源有通用寄存器组、外部直接输入 D_A 和 D_B 、乘商寄存器 MQ 。

D_i 、 D_o 分别是位片的数据输入、输出端。虽然每片只有 4 位,但将若干位拼接起来,再加上微程序控制器芯片,就可方便地构成中央处理器 CPU。

4.4 教材习题解答

1. 证明在全加器里,进位传递函数 $P=A_i+B_i=A_i\oplus B_i$ 。

解: 并行加法器中的每一个全加器都有一个从低位送来的进位和一个传送给较高位的进位。进位表达式为

$$C_i = A_i B_i + (A_i \oplus B_i) C_{i-1}$$

欲证明 $P_i = A_i + B_i = A_i \oplus B_i$, 也就是要证明 $C_i = A_i B_i + (A_i \oplus B_i) C_{i-1} = A_i B_i + (A_i + B_i) C_{i-1}$

用卡诺图法,图 4-10(a)和 4-10(b)分别是两个逻辑表达式的卡诺图。两个卡诺图相同,两个逻辑表达式就相等,则进位传递函数的两种形式相等。

2. 某加法器采用组内并行、组间并行的进位链,4 位一组,写出进位信号 C_6 的逻辑表达式。

解: 最低一组的进位输出 $C_4 = G_4^* + P_4^* C_0$

其中:

$$G_4^* = G_4 + P_4 G_3 + P_4 P_3 G_2 + P_4 P_3 P_2 G_1$$

$$P_4^* = P_4 P_3 P_2 P_1$$

$$C_5 = G_5 + P_5 C_4$$

所以

$$C_6 = G_6 + P_6 C_5 = G_6 + P_6 G_5 + P_6 P_5 C_4$$

3. 设计一个 9 位先行进位加法器,每 3 位为一组,采用两级先行进位线路。

$$C_1 = G_1 + P C_0$$

解:

$$C_2 = G_2 + P_2 G_1 + P_2 P_1 C_0$$

$$C_3 = G_3 + P_3 G_2 + P_3 P_2 G_1 + P_3 P_2 P_1 C_0$$

设: $G_3^* = G_3 + P_3 G_2 + P_3 P_2 G_1$, $P_3^* = P_3 P_2 P_1$

$$C_3 = G_3^* + P_3^* C_0$$

则有:

$$C_6 = G_6^* + P_6^* G_3^* + P_6^* P_3^* C_0$$

$$C_9 = G_9^* + P_9^* G_6^* + P_9^* P_6^* G_3^* + P_9^* P_6^* P_3^* C_0$$

9 位先行进位加法器如图 4-11 所示。

4. 已知 X 和 Y , 试用它们的变形补码计算出 $X+Y$, 并指出结果是否溢出。

(1) $X=0.11011, Y=0.11111$

(2) $X=0.11011, Y=-0.10101$

(3) $X=-0.10110, Y=-0.00001$

(4) $X = -0.11011, Y = 0.11110$

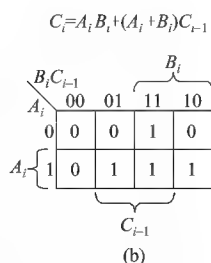
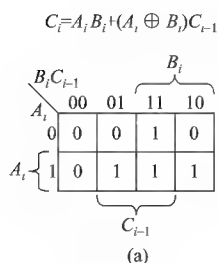


图 4 10 全加器的卡诺图

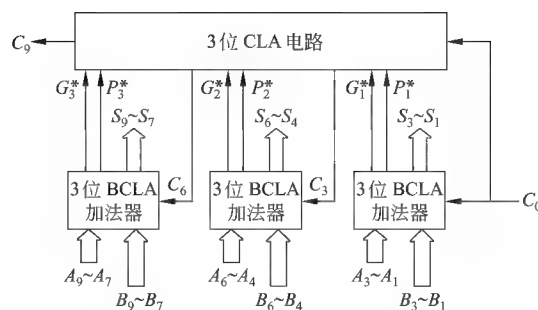


图 4 11 9 位先行进位加法器

解: (1) $[X]_{\text{补}} = 0.11011, [Y]_{\text{补}} = 0.11111$

$$\begin{array}{r} 00.11011 \\ + 00.11111 \\ \hline 01.11010 \end{array} \quad \begin{array}{l} [X]_{\text{补}} \\ [Y]_{\text{补}} \\ [X+Y]_{\text{补}} \end{array} \quad \text{结果正溢}$$

(2) $[X]_{\text{补}} = 0.11011, [Y]_{\text{补}} = 1.01011$

$$\begin{array}{r} 00.11011 \\ + 11.01011 \\ \hline 00.00110 \end{array} \quad \begin{array}{l} [X]_{\text{补}} \\ [Y]_{\text{补}} \\ [X+Y]_{\text{补}} \end{array}$$

$X+Y = 0.00110$

(3) $[X]_{\text{补}} = 1.01010, [Y]_{\text{补}} = 1.11111$

$$\begin{array}{r} 11.01010 \\ + 11.11111 \\ \hline 11.01001 \end{array} \quad \begin{array}{l} [X]_{\text{补}} \\ [Y]_{\text{补}} \\ [X+Y]_{\text{补}} \end{array}$$

$X+Y = -0.10111$

(4) $[X]_{\text{补}} = 1.00101, [Y]_{\text{补}} = 0.11110$

$$\begin{array}{r} 11.00101 \\ + 00.11110 \\ \hline 00.00011 \end{array} \quad \begin{array}{l} [X]_{\text{补}} \\ [Y]_{\text{补}} \\ [X+Y]_{\text{补}} \end{array}$$

$X+Y = 0.00011$

5. 已知 X 和 Y , 试用它们的变形补码计算出 $X-Y$, 并指出结果是否溢出。

(1) $X = 0.11011, Y = -0.11111$

(2) $X = 0.10111, Y = 0.11011$

(3) $X = 0.11011, Y = -0.10011$

(4) $X = -0.10110, Y = -0.00001$

解: (1) $[X]_{\text{补}} = 0.11011, [Y]_{\text{补}} = 1.00001, [-Y]_{\text{补}} = 0.11111$

$$\begin{array}{r} 00.11011 \quad [X]_{\text{补}} \\ + 00.11111 \quad [-Y]_{\text{补}} \\ \hline 01.11010 \quad [X-Y]_{\text{补}} \quad \text{结果正溢} \end{array}$$

(2) $[X]_{\text{补}} = 0.10111, [Y]_{\text{补}} = 0.11011, [-Y]_{\text{补}} = 1.00101$

$$\begin{array}{r} 00.10111 \quad [X]_{\text{补}} \\ + 11.00101 \quad [-Y]_{\text{补}} \\ \hline 11.11100 \quad [X-Y]_{\text{补}} \end{array}$$

$$X - Y = -0.00100$$

(3) $[X]_{\text{补}} = 0.11011, [Y]_{\text{补}} = 1.01101, [-Y]_{\text{补}} = 0.10011$

$$\begin{array}{r} 00.11011 \quad [X]_{\text{补}} \\ + 00.10011 \quad [-Y]_{\text{补}} \\ \hline 01.01110 \quad [X-Y]_{\text{补}} \quad \text{结果正溢} \end{array}$$

(4) $[X]_{\text{补}} = 1.01010, [Y]_{\text{补}} = 1.11111, [-Y]_{\text{补}} = 0.00001$

$$\begin{array}{r} 11.01010 \quad [X]_{\text{补}} \\ + 00.00001 \quad [-Y]_{\text{补}} \\ \hline 11.01011 \quad [X-Y]_{\text{补}} \end{array}$$

$$X - Y = -0.10101$$

6. 已知: $X = 0.1011, Y = -0.0101$

求: $\left[\frac{1}{2}X\right]_{\text{补}}, \left[\frac{1}{4}X\right]_{\text{补}}, [-X]_{\text{补}}, \left[\frac{1}{2}Y\right]_{\text{补}}, \left[\frac{1}{4}Y\right]_{\text{补}}, [-Y]_{\text{补}}$ 。

解: $[X]_{\text{补}} = 0.1011$

$$\left[\frac{1}{2}X\right]_{\text{补}} = 0.0101, \left[\frac{1}{4}X\right]_{\text{补}} = 0.0010, [-X]_{\text{补}} = 1.0101$$

$$[Y]_{\text{补}} = 1.1011$$

$$\left[\frac{1}{2}Y\right]_{\text{补}} = 1.1101, \left[\frac{1}{4}Y\right]_{\text{补}} = 1.1110, [-Y]_{\text{补}} = 0.0101$$

7. 设下列数据长 8 位, 包括 1 位符号位, 采用补码表示, 分别写出每个数据右移或左移 2 位之后的结果。

(1) 0.1100100

(2) 1.0011001

(3) 1.1100110

(4) 1.0000111

解: (1) $[X]_{\text{补}} = 0.1100100$

$$\left[\frac{1}{4}X\right]_{\text{补}} = 0.0011001, [4X]_{\text{补}} = 0.0010000$$

(2) $[X]_{\text{补}} = 1.0011001$

$\left[\frac{1}{4}X\right]_{\text{补}} = 1.1100110, [4X]_{\text{补}} = 1.1100100$

(3) 1.1100110

$\left[\frac{1}{4}X\right]_{\text{补}} = 1.1111001, [4X]_{\text{补}} = 1.0011000$

(4) 1.0000111

$\left[\frac{1}{4}X\right]_{\text{补}} = 1.1100001, [4X]_{\text{补}} = 1.0011100$

8. 分别用原码乘法和补码乘法计算 $X \times Y$ 。

(1) $X = 0.11011, Y = -0.11111$

(2) $X = -0.11010, Y = -0.01110$

解: (1) 原码乘法:

$|X| = 0.11011 \rightarrow B, |Y| = 0.11111 \rightarrow C, 0 \rightarrow A$

	A	C	说明
	00.00000	0.11111	
$+ X $	00.11011		$C_5=1, + X $
	00.11011		
\rightarrow	00.01101	10.11111	部分积右移1位
$+ X $	00.11011		$C_5=1, + X $
	01.01000		
\rightarrow	00.10100	01.11111	部分积右移1位
$+ X $	00.11011		$C_5=1, + X $
	01.01111		
\rightarrow	00.10111	10.11111	部分积右移1位
$+ X $	00.11011		$C_5=1, + X $
	01.10010		
\rightarrow	00.11001	01.01111	部分积右移1位
$+ X $	00.11011		$C_5=1, + X $
	01.10100		
\rightarrow	00.11010	00.10111	部分积右移1位

所以

$|X \times Y| = 0.1101000101$

$X \times Y = -0.1101000101$

补码乘法：

$[X]_{\text{补}}=0.11011 \rightarrow B$, $[Y]_{\text{补}}=1.00001 \rightarrow C$, $0 \rightarrow A$

$[-X]_{\text{补}}=1.00101$

A	C	附加位	说明
00.00000	1.0000 <u>1</u> 0	↓	
+ $[-X]_{\text{补}}$ 1.100101			$C_5 C_6=10$, $+[-X]_{\text{补}}$
1.100101			
→ 1.110010	1.1000 <u>0</u> 1		部分积右移 1 位
+ $[X]_{\text{补}}$ 0.011011			$C_5 C_6=01$, $+ [X]_{\text{补}}$
00.01101			
→ 00.00110	1.1100 <u>0</u> 0		部分积右移 1 位
+0 00.00000			$C_5 C_6=00$, +0
00.00110			
→ 00.00011	0.1110 <u>0</u> 0		部分积右移 1 位
+0 00.00000			$C_5 C_6=00$, +0
00.00011			
→ 00.00001	1.0111 <u>0</u> 0		部分积右移 1 位
+0 00.00000			$C_5 C_6=00$, +0
00.00001			
→ 00.00000	1.1011 <u>1</u> 0		部分积右移 1 位
+ $[-X]_{\text{补}}$ 1.100101			$C_5 C_6=10$, $+[-X]_{\text{补}}$
1.100101			

所以

$$[X \times Y]_{\text{补}} = 1.0010111011$$

$$X \times Y = -0.1101000101$$

(2) $X \times Y = 0.0101101100$, 过程略。

9. 根据补码两位乘法规则推导出补码 3 位乘法的规则。

解：先根据补码 1 位乘法推出补码 2 位乘法规则,再根据补码 2 位乘法推出补码 3 位乘法规则。

$$[Z']_{\text{补}} = 2^{-1} \{ [Z]_{\text{补}} + (Y_{i+1} - Y_i) [X]_{\text{补}} \}$$

$$[Z'']_{\text{补}} = 2^{-1} \{ [Z']_{\text{补}} + (Y_i - Y_{i-1}) [X]_{\text{补}} \}$$

$$= 2^{-2} \{ [Z]_{\text{补}} + (Y_{i+1} + Y_i - 2Y_{i-1}) [X]_{\text{补}} \}$$

$$[Z''']_{\text{补}} = 2^{-1} \{ [Z'']_{\text{补}} + (Y_{i-1} - Y_{i-2}) [X]_{\text{补}} \}$$

$$= 2^{-1} \{ 2^{-2} \{ [Z]_{\text{补}} + (Y_{i+1} + Y_i - 2Y_{i-1}) [X]_{\text{补}} \} + (Y_{i-1} - Y_{i-2}) [X]_{\text{补}} \}$$

$$= 2^{-3} \{ [Z]_{\text{补}} + (Y_{i+1} + Y_i - 2Y_{i-1}) \times [X]_{\text{补}} + 2^2 \times (Y_{i-1} - Y_{i-2}) [X]_{\text{补}} \}$$

$$= 2^{-3} \{ [Z]_{\text{补}} + (Y_{i+1} + Y_i + 2Y_{i-1} - 4Y_{i-2}) \times [X]_{\text{补}} \}$$

10. 分别用原码和补码加减交替法计算 $X \div Y$ 。

- (1) $X=0.10101, Y=0.11011$
 (2) $X=-0.10101, Y=0.11011$
 (3) $X=0.10001, Y=-0.10110$
 (4) $X=-0.10110, Y=-0.11011$

解: (1) 原码除法:

$|X|=0.10101 \rightarrow A, |Y|=0.11011 \rightarrow B, 0 \rightarrow C$

$[|Y|]_{\text{变补}}=1.00101$

A	C	说明
00.10101	0.00000	
$+ [Y]_{\text{变补}} \quad 11.00101$		$- Y $
11.11010	0.00000	部分余数为负, 商0
$\leftarrow 11.110100$		左移1位
$+ Y \quad 00.11011$		$+ Y $
00.01111	0.00001	部分余数为正, 商1
$\leftarrow 00.11110$		左移1位
$+ [Y]_{\text{变补}} \quad 11.00101$		$- Y $
00.00011	0.00011	部分余数为正, 商1
$\leftarrow 00.00110$		左移1位
$+ [Y]_{\text{变补}} \quad 11.00101$		$- Y $
11.01011	0.00110	部分余数为负, 商0
$\leftarrow 10.10110$		左移1位
$+ Y \quad 00.11011$		$+ Y $
11.10001	0.01100	部分余数为负, 商0
$\leftarrow 11.00010$		左移1位
$+ Y \quad 00.11011$		$+ Y $
11.11101	0.11000	部分余数为负, 商0
$+ Y \quad 00.11011$		最后一次恢复余数, $+ Y $
00.11000		

因为

$$Q_s = X_s \oplus Y_s = 0 \oplus 0 = 0$$

所以

$$\frac{X}{Y} = 0.11000 + \frac{0.11000 \times 2^{-5}}{0.11011}$$

补码除法：

$[X]_{\text{补}}=0.10101 \rightarrow A$, $[Y]_{\text{补}}=0.11011 \rightarrow B$, $0 \rightarrow C$

$[-Y]_{\text{补}}=1.00101$

A	C	说明
00.10101	0.00000	
$+ [-Y]_{\text{补}}$ 11.100101		$[X]_{\text{补}}$ 、 $[Y]_{\text{补}}$ 同号, $+ [-Y]_{\text{补}}$
11.11010	0.00000	$[r_i]_{\text{补}}$ 、 $[Y]_{\text{补}}$ 异号, 商0
← 11.10100		左移1位
$+ [Y]_{\text{补}}$ 00.11011		$+ [Y]_{\text{补}}$
00.01111	0.00001	$[r_i]_{\text{补}}$ 、 $[Y]_{\text{补}}$ 同号, 商1
← 00.11110		左移1位
$+ [-Y]_{\text{补}}$ 11.100101		$+ [-Y]_{\text{补}}$
00.00011	0.00011	$[r_i]_{\text{补}}$ 、 $[Y]_{\text{补}}$ 同号, 商1
← 00.00110		左移1位
$+ [-Y]_{\text{补}}$ 11.100101		$+ [-Y]_{\text{补}}$
11.01011	0.00110	$[r_i]_{\text{补}}$ 、 $[Y]_{\text{补}}$ 异号, 商0
← 10.10110		左移1位
$+ [Y]_{\text{补}}$ 00.11011		$+ [Y]_{\text{补}}$
11.10001	0.01100	$[r_i]_{\text{补}}$ 、 $[Y]_{\text{补}}$ 异号, 商0
← 11.00010		左移1位
$+ [Y]_{\text{补}}$ 00.11011		$+ [Y]_{\text{补}}$
11.11101	0.11001	末位恒置1

所以 $\left[\frac{X}{Y} \right]_{\text{补}} = 0.11001 + \frac{1.11101 \times 2^{-5}}{0.11011}$

$$\frac{X}{Y} = 0.11001 + \frac{-0.00011 \times 2^{-5}}{0.11011}$$

(2) 中间过程略。原码除法: $X \div Y = -0.11000 + \frac{0.11000 \times 2^{-5}}{0.11011}$

补码除法: $X \div Y = -0.11001 + \frac{0.00011 \times 2^{-5}}{0.11011}$

(3) 中间过程略。原码除法: $X \div Y = -0.11000 + \frac{0.10000 \times 2^{-5}}{0.10110}$

补码除法: $X \div Y = -0.11001 + \frac{0.00101 \times 2^{-5}}{0.10110}$

(4) 中间过程略。原码除法： $X \div Y = 0.11010 + \frac{0.00010 \times 2^{-5}}{0.11011}$

补码除法： $X \div Y = 0.11011 - \frac{0.11001 \times 2^{-5}}{0.11011}$

11. 设浮点数的阶码和尾数部分均用补码表示,按照浮点数的运算规则,计算下列各题:

(1) $X = 2^{101} \times (-0.100010), Y = 2^{100} \times (-0.111110)$

(2) $X = 2^{-101} \times 0.101100, Y = 2^{-100} \times (-0.101000)$

(3) $X = 2^{-011} \times 0.101100, Y = 2^{-001} \times (-0.111100)$

注:此题中阶码用二进制表示。

求: $X+Y, X-Y$ 。

解: (1) $X = 2^{101} \times (-0.100010), Y = 2^{100} \times (-0.111110)$

$$[X]_{\text{浮}} = 0101; 1.011110$$

$$[Y]_{\text{浮}} = 0100; 1.000010$$

对阶:小阶向大阶看齐, $\Delta E = E_A - E_B = 1$,

$$[Y]_{\text{浮}}' = 0101; 1.100001$$

对阶之后,尾数相加和相减。

相加:

$$\begin{array}{r} 11.011110 \\ + 11.100001 \\ \hline 10.111111 \end{array}$$

需右规一次, $[X+Y]_{\text{浮}} = 0110; 1.011111$

所以

$$X+Y = 2^{110} \times (-0.100001)$$

相减:

$$\begin{array}{r} 11.011110 \\ + 00.011111 \\ \hline 11.111101 \end{array}$$

需左规4次, $[X-Y]_{\text{浮}} = 0001; 1.010000$

所以

$$X-Y = 2^{001} \times (-0.110000)$$

(2) $X = 2^{-101} \times 0.101100, Y = 2^{-100} \times (-0.101000)$

$$[X]_{\text{浮}} = 1011; 0.101100$$

$$[Y]_{\text{浮}} = 1100; 1.011000$$

对阶:小阶向大阶看齐。 $\Delta E = E_A - E_B = -1$

$$[X]_{\text{浮}}' = 1100; 0.010110$$

对阶之后,尾数相加和相减。

$$\begin{array}{r} \text{相加:} \qquad \qquad \qquad 00.010110 \\ \qquad \qquad \qquad + 11.011000 \\ \hline \qquad \qquad \qquad 11.101110 \end{array}$$

需左规一次, $[X+Y]_{\text{浮}} = 1011; 1.011100$

所以 $X+Y = 2^{-101} \times (-0.100100)$

$$\begin{array}{r} \text{相减:} \qquad \qquad \qquad 00.010110 \\ \qquad \qquad \qquad + 00.101000 \\ \hline \qquad \qquad \qquad 00.111110 \end{array}$$

所以 $X-Y = 2^{-100} \times 0.111110$

(3) $X = 2^{-011} \times 0.101100, Y = 2^{-001} \times (-0.111100)$

$$[X]_{\text{浮}} = 1101; 0.101100$$

$$[Y]_{\text{浮}} = 1111; 1.000100$$

对阶: 小阶向大阶看齐。 $\Delta E = E_A - E_B = -2$

$$[X]_{\text{浮}}' = 1111; 0.001011$$

对阶之后, 尾数相加和相减。

$$\begin{array}{r} \text{相加:} \qquad \qquad \qquad 00.001011 \\ \qquad \qquad \qquad + 11.000100 \\ \hline \qquad \qquad \qquad 11.001111 \end{array}$$

所以 $X+Y = 2^{-001} \times (-0.110001)$

$$\begin{array}{r} \text{相减:} \qquad \qquad \qquad 00.001011 \\ \qquad \qquad \qquad + 00.111100 \\ \hline \qquad \qquad \qquad 01.000111 \end{array}$$

需右规一次, $[X-Y]_{\text{浮}} = 0000; 0.100011$

所以 $X-Y = 2^{-000} \times 0.100011$

12. 设浮点数的阶码和尾数部分均用补码表示, 按照浮点数的运算规则, 计算下列各题:

$$(1) X = 2^3 \times \frac{13}{16}, Y = 2^4 \times \left| -\frac{9}{16} \right|$$

求: $X \times Y$ 。

$$(2) X = 2^3 \times \left| -\frac{13}{16} \right|, Y = 2^5 \times \left| \frac{15}{16} \right|$$

求: $X \div Y$ 。

$$\text{解: (1) } X = 2^3 \times \frac{13}{16}, Y = 2^4 \times \left| -\frac{9}{16} \right|$$

阶码相加： $E_A + E_B = 3 + 4 = 7$

尾数相乘：由补码乘法规则求得： -0.01110101

结果规格化：左规一次， $X \times Y = -0.11101010 \times 2^6$

$$(2) X = 2^3 \times \left(-\frac{13}{16}\right), Y = 2^5 \times \left(\frac{15}{16}\right)$$

尾数调整：因为 $|X_{\text{尾数}}| \leq |Y_{\text{尾数}}|$ ，所以无需尾数调整。

阶码相减： $E_A - E_B = 3 - 5 = -2$

尾数相除：由补码除法规则求得： $-0.1101 + \frac{-0.1101 \times 2^4}{0.1111}$

$$X \div Y = \left| -0.1101 + \frac{-0.1101 \times 2^4}{0.1111} \right| \times 2^{-2}$$

13. 用流程图描述浮点除法运算的算法步骤。

解：浮点除法运算的算法流程图如图 4-12 所示。

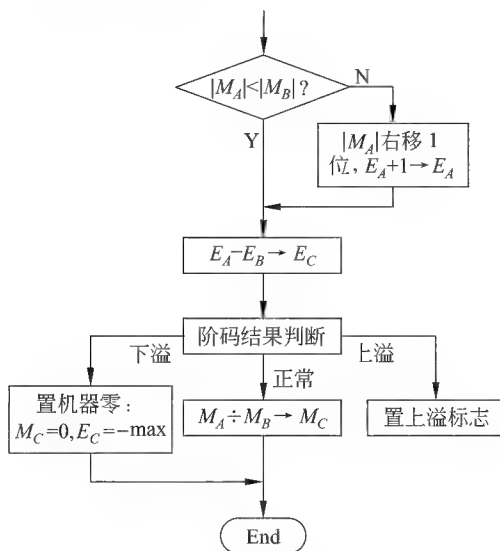


图 4-12 浮点除法运算流程图

14. 设计一个 1 位 5421 码加法器。

解：设 1 位被加数为 $A_4 A_3 A_2 A_1$ ，加数为 $B_4 B_3 B_2 B_1$ 。5421 码的校正关系如表 4-4 所示。

表 4 4 5421 码的校正关系

	十进制数	5421 码 $C_4 S_4 S_3 S_2 S_1$	校正前的二进制数 $C'_4 S'_4 S'_3 S'_2 S'_1$	校正关系
①	0	0 0 0 0 0	0 0 0 0 0	不校正
	1	0 0 0 0 1	0 0 0 0 1	
	2	0 0 0 1 0	0 0 0 1 0	
	3	0 0 0 1 1	0 0 0 1 1	
	4	0 0 1 0 0	0 0 1 0 0	
②	5	0 1 0 0 0	0 0 1 0 1	若 $A < 5, B < 5$, 则+3 校正
	6	0 1 0 0 1	0 0 1 1 0	
	7	0 1 0 1 0	0 0 1 1 1	
	8	0 1 0 1 1	0 1 0 0 0	
	9	0 1 1 0 0	0 1 0 0 1	
③	10	1 0 0 0 0	0 1 1 0 1	若 A 或 $B \geq 5, B$ 或 $A < 5$, 则+3 校正
	11	1 0 0 0 1	0 1 1 1 0	
	12	1 0 0 1 0	0 1 1 1 1	
	13	1 0 0 1 1	1 0 0 0 0	
	14	1 0 1 0 0	1 0 1 0 1	
④	15	1 1 0 0 0	1 0 1 0 1	若 $A \geq 5, B \geq 5$, 则+3 校正
	16	1 1 0 0 1	1 0 1 1 0	
	17	1 1 0 1 0	1 0 1 1 1	
	18	1 1 0 1 1	1 1 0 0 0	
	19	1 1 1 0 0	1 1 0 0 1	

① 和在 0~4 范围内,不用校正,结果正确。

② 和在 6~9 范围内,当 $A < 5, B < 5$,需+3 校正,而当 $A < 5, B \geq 5$ 或 $A \geq 5, B < 5$ 时,不需校正。故校正函数为:

$$\overline{A_4} \overline{B_4} (S'_4 + S'_3 S'_2 + S'_3 S'_1)$$

③ 和在 10~14 范围内,当 $A < 5, B \geq 5$,或 $A \geq 5, B < 5$,需+3 校正,而当 $A \geq 5, B \geq 5$ 时,不需校正。故校正函数为:

$$(A_4 \oplus B_4)(C'_4 + S'_3 S'_2 + S'_3 S'_1)$$

④ 和在 16~19 范围内($A \geq 5, B \geq 5$),一定+3 校正。

$$A_4 B_4 (S'_4 + S'_3 S'_2 + S'_3 S'_1)$$

将 3 部分校正函数统一考虑并化简,得:

$$\begin{aligned}\text{校正函数} &= S'_3 S'_2 + S'_3 S'_1 + (A_4 \oplus B_4) S'_4 + (A_4 \oplus B_4) C'_4 \\ &= S'_3 S'_2 + S'_3 S'_1 + C'_3\end{aligned}$$

15. 某机利用二进制的加法器进行 8421 码的十进制运算, 采用的方法是:

① 对某一操作数预加 6 后, 与另一操作数一起进入二进制加法器;

② 有进位产生时, 直接得到和的

8421 码;

③ 没有进位时, 反减 6 再得到和的

8421 码。

试求 +6、-6 的校正逻辑。

解: 设某一操作数为 $A_4 A_3 A_2 A_1$, +6 的校正后的操作数为 $A'_4 A'_3 A'_2 A'_1$; 设校正前和为 $S'_4 S'_3 S'_2 S'_1$, 进位为 C'_4 , 若 $C'_4 = 1$, 即为正确和 $S_4 S_3 S_2 S_1$, 若 $C'_4 = 0$, -6 校正 (+1010), 其加法器逻辑图如图 4-13 所示。

16. 用 74181 和 74182 芯片构成一个 64 位的 ALU, 采用多级分组并行进位链 (要求速度尽可能快)。

解: 共需要 16 片 74181, 5 片 74182

组成二级先行进位的 64 位 ALU。如图 4-14 所示。

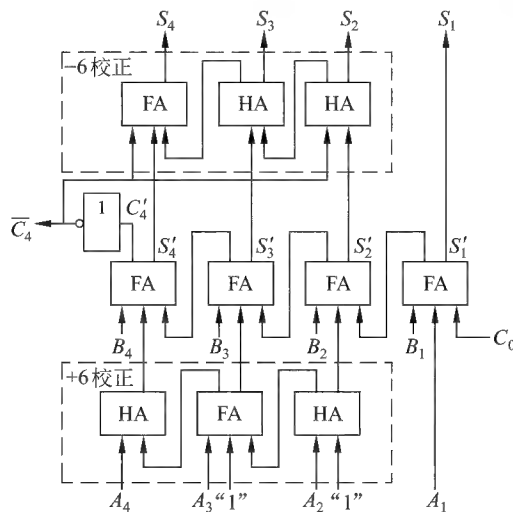


图 4-13 8421 码加法器逻辑图

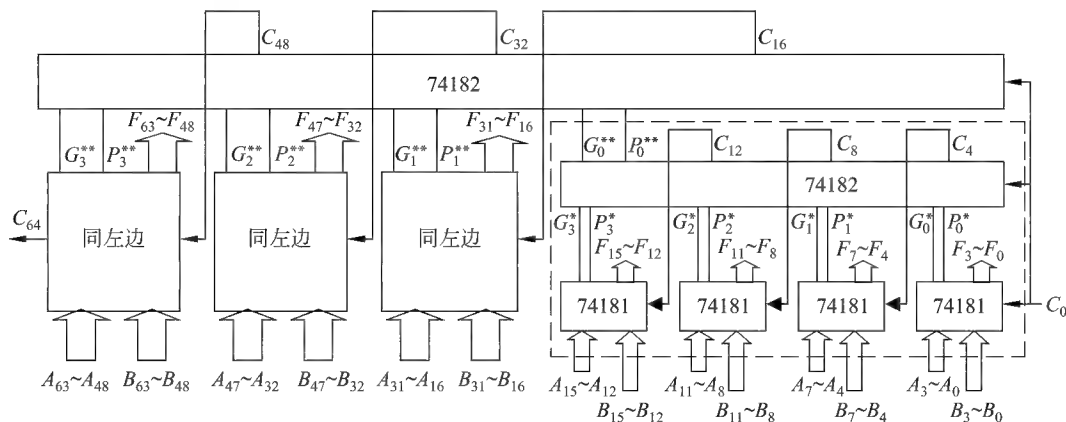


图 4-14 三级先行进位的 64 位 ALU

第 5 章

存储系统和结构

5.1 基本内容要求

存储系统是由几个容量、速度和价格各不相同的存储器构成的系统,设计一个容量大、速度快、成本低的存储系统是计算机发展的一个重要课题。本章重点讨论主存储器的工作原理、组成方式以及运用半导体存储芯片组成主存储器的一般原则和方法,此外还介绍高速缓冲存储器和虚拟存储器的基本原理。

学习要求

- ◆ 了解存储器的各种分类方法
- ◆ 了解存储系统的两个层次(Cache—主存层次,主—辅存层次)
- ◆ 了解主存储器的基本结构
- ◆ 理解主存储器的有关术语(如位、存储字、存储单元、存储体等)
- ◆ 理解主存储器的主要技术指标
- ◆ 掌握字节编址存储器的各种访问方法,将不同长度的数据按要求存放在存储器中
- ◆ 了解半导体随机存储器(静态 RAM 和动态 RAM)不同的基本存储原理
- ◆ 理解动态 RAM 3 种不同刷新方式的特点
- ◆ 了解 RAM 芯片的基本结构
- ◆ 理解各种不同 ROM 的特点
- ◆ 理解主存储器中包括 RAM 和 ROM 两种形式
- ◆ 掌握主存储器容量的各种扩展方法,使用若干存储芯片构成存储器
- ◆ 掌握存储芯片的地址分配和片选信号的产生
- ◆ 理解主存储器和 CPU 的软连接(读写操作)
- ◆ 理解主存的奇偶校验和 ECC
- ◆ 理解 PC 系列微机的存储器接口
- ◆ 了解提高 RAM 芯片速度的技术

- ◆ 了解并行交叉存储技术
- ◆ 了解 Cache 的特点
- ◆ 了解虚拟存储器的概念

5.2 误点疑点解惑

1. 存储系统和存储器

在同一台计算机中,有各种工作速度、存储容量、访问方式、用途等均不相同的存储器,这些存储器构成一个层次结构,如图 5-1 所示。从上到下,各种存储器的存储容量越来越大,每位的价格越来越便宜,但存取周期越来越长。

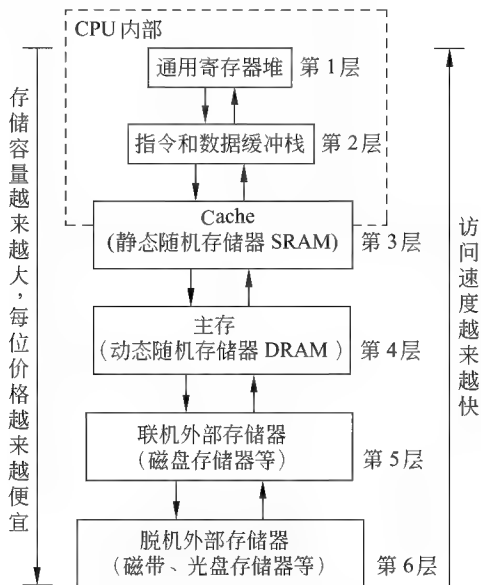


图 5 1 存储器的层次结构

需要提醒学生注意的是,并非将各种用途不同的存储器放在一起就构成了一个存储系统。存储系统是指两个或两个以上速度、容量和价格各不相同的存储器用硬件、软件、硬件与软件相结合的方法连接起来的一个系统。这个系统对应用程序员透明,可以把它看作是一个“存储器”,其速度接近速度最快的那个存储器,存储容量与容量最大的那个存储器相等或接近,单位容量的价格接近最便宜的那个存储器。

所以说,存储系统和存储器是两个完全不同的概念。如果在一台计算机中只有存储器,甚至有多种存储器,但没有存储系统,这台计算机的性能将会是很差的,这些存储器的

性能也不可能得到充分地发挥。

2. 主存储器组织

主存储器的核心是存储体,程序和数据都存放在存储体中。存储体是由若干存储单元组成的,存储单元的编号称为地址,地址和存储单元之间有一对一的对应关系。这就像一座大楼有许多房间,而每个房间都有其惟一的房间号一样。

位是二进制数的最小单位,是半导体存储器的基本记忆单元。存储字由若干二进制位组成,可以作为一个整体存入或取出。一个存储单元可能存放一个字,也可能存放一个字节,这是由计算机的结构确定的。对于字编址的计算机,最小寻址单位是一个字,相邻的存储单元地址指向相邻的存储字;对于字节编址的计算机,最小寻址单位是一个字节,相邻的存储单元地址指向相邻的存储字节。所以,存储单元是 CPU 对主存可访问操作的最小存储单位,根据存储单元的地址可以找到相应存储单元的内容。

3. 字节编址计算机的大端方案和小端方案

每个字中的字节地址可以从左到右或者从右到左编排,前者称为大端方案,后者称为小端方案。Intel 80x86 是采用小端方案的机器,IBM 370、Motorola 680x0 和大多数 RISC 机器则采用大端方案,Power PC 是一个既支持大端方案又支持小端方案的机器。图 5-2(a)描述的是使用大端方案的 32 位计算机的一段主存,图 5-2(b)描述的是使用小端方案的 32 位计算机的一段主存。图中每个存储单元上的数字表示字节地址。

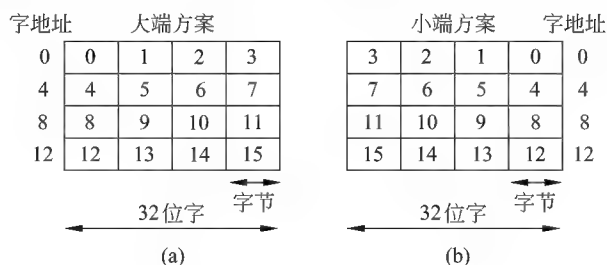


图 5 2 大端和小端方案

大端方案和小端方案在存放 ASCII 码字符串和 BCD 码数据的顺序是相反的;但是必须指出的是,不管是上述哪个系统用来表示一个 32 位整数,两个方案是一致的。比如 6,都是在最右边的(最低位)3 位上存放有 110,前面 29 位都是 0。也就是说,在大端方案中,110 这 3 位应该放在字节 3(或 7、11 等)中,而在小端方案中,110 这 3 位应该放在字节 0 或 4、8 等中。

如果计算机只用来存放整数,也不会有任何问题,然而,许多应用中要存放的是整数、字符串和其他数据类型的混合结构,这就可以出现混乱。这个问题的解决方法是在每个数据项前面加上一个头来描述其后的数据类型和数据长度,使接收方可对数据进行必要

的转换。

4. 主存储器的存储容量和存取速度

描述主存储器性能的主要技术指标是存储容量和存取速度。

(1) 存储容量

存储容量是指整个主存储器所能存放的二进制信息的总位数,可以这样定义:

$$S_M = W \times L$$

其中, W 为存储字数, L 为存储器字长。

例如,某存储器字长 16 位,共有 1024 存储字,那么该存储器的容量为:

$$1024 \times 16 = 16\,384 (\text{位}) = 2048 (\text{字节})$$

需要注意的是,在存储器中常称 16 384 位为 16K 位或 2K 字节,这是因为存储器中的 1K 不是 1000,而是 $1024(2^{10})$ 。1M 不是 1 000 000,而是 $1\,048\,576(2^{20})$ 。

(2) 存取速度

主存的存取速度通常由存取时间 T_a 、存取周期 T_m 和主存带宽 B_m 等参数来描述。

存取时间 T_a 是执行一次读操作或写操作的时间,即从地址传送给主存开始到数据能够被使用为止所用的时间间隔。存取周期 T_m 是指两次连续的存储器操作(如两次连续的读操作)之间所需要的最小时间间隔。一般情况下, $T_m > T_a$,对于破坏性读出的存储器, $T_m = 2T_a$ 。

主存带宽 B_m 是指连续访问主存时主存所能提供的数据传送率。例如,对于 SDRAM 而言,若工作频率为 100MHz,其数据传输率可以达到 800MBps ($100 \times 64 \div 8 = 800$);若工作频率为 133MHz,其数据传输率可以达到 1.06GBps ($133 \times 64 \div 8 = 1064$)。对于 DDR SDRAM 而言,由于在同一个时钟的上升沿和下降沿都能传输数据,所以工作频率在 200MHz 时,数据传输率可以达到 3.2GBps ($200 \times 64 \times 2 \div 8 = 3200$)。

5. 边界对齐的数据存放方法

在采用字节编址的存储器中,数据有 3 种不同的存放方法,其中边界对齐的存放方法是最有效的方法。边界对齐的数据存放方式对数据的存放位置有下列要求:

8 位数据,占有 1 个存储单元,其地址为 $\times \cdots \times \times \times \times$ (任意)

16 位数据,占用 2 个存储单元,存放数据的起始地址为 $\times \cdots \times \times \times 0$ (2 的整倍数)

32 位数据,占用 4 个存储单元,存放数据的起始地址为 $\times \cdots \times \times 0 0$ (4 的整倍数)

64 位数据,占用 8 个存储单元,存放数据的起始地址为 $\times \cdots \times 0 0 0$ (8 的整倍数)

例 5-1 某机字长 32 位,主存储器按字节编址,现有 4 种不同长度的数据(字节、半字、单字、双字),请采用一种既节省存储空间,又能保证任何长度的数据都在单个存取周期内完成读写的方法,将一批数据顺序地存入主存,画出主存中数据的存放示意图。

这批数据一共有 10 个,它们依次为字节、半字、双字、单字、字节、单字、双字、半字、单字、字节。

解：根据题干可以知道 4 种长度的数据分别为：字节数据 8 位，半字数据 16 位，单字数据 32 位，双字数据 64 位。因为要保证任何长度的数据都在单个存取周期内完成读写，所以该机的存储字长应为 64 位。要特别注意的是，在本例中数据字长（32 位）和存储字长（64 位）是不同的。

题干中要求采用一种既节省存储空间，又能保证任何长度的数据都在单个存取周期内完成读写的方法来顺序存入一批数据，所以只能选用边界对齐的存放方法，双字数据从字节地址为 8 的整倍数的地方开始存放，单字数据从字节地址为 4 的整倍数的地方开始存放，半字地址从字节地址为 2 的整倍数的地方开始存放。主存中数据的存放示意图如图 5-3 所示。

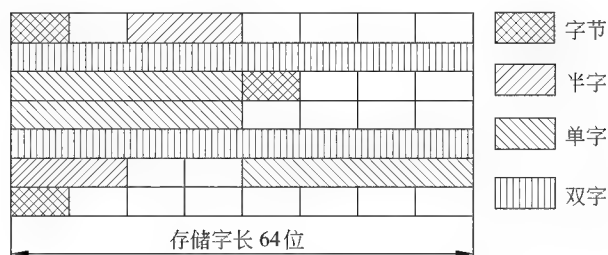


图 5-3 主存中数据的存放示意图

6. 动态随机存储器的刷新

动态随机存储器 (DRAM) 利用栅极电容来存储信息，电容上的电荷会随着时间推移而泄漏掉，必须定时刷新。

在讨论刷新问题时，首先要搞清楚刷新和重写（再生）这两个完全不同的概念。重写是随机的，某个存储单元只有在破坏性读出之后才需要重写；而刷新是定时的，即使许多记忆单元长期未被访问，若不及时补充电荷的话，信息也会丢失。重写是按存储单元进行的，破坏性地读出了哪个单元就只对这个单元重写，而不需要涉及其他的存储单元；而刷新则不论某个单元是否被读出均需要进行，所以是以存储体矩阵中的一行为单位进行的。

刷新的方式主要有集中式、分散式和异步式 3 种。其中异步刷新方式是一种比较实用的刷新方式，它的死区比较小，且刷新次数比较少。

刷新通常是一行一行地进行的，每一行中各记忆单元同时被刷新，故仅需要行地址，不需要列地址。由刷新控制电路中的刷新计数器产生行地址，刷新操作类似于读出操作，但仅有 $\overline{\text{RAS}}$ 信号。

整个存储器中的所有芯片同时被刷新。在考虑刷新问题时，应当从单个芯片的存储容量着手，而不是从整个存储器的容量着手。这是学生比较容易犯的一个错误，需要特别强调。

7. 各类半导体存储芯片的特点

半导体存储器包括半导体随机存储器(RAM)和半导体只读存储器(ROM)。RAM多用MOS型电路组成,MOS RAM按电路结构不同又可以分为静态RAM(SRAM)和动态RAM(DRAM)。

RAM是可读、可写的存储器,CPU可以对RAM的内容随机地读写访问。SRAM的存取速度快,但集成度低,功耗也较大,所以一般用来组成高速缓冲存储器和小容量主存系统;DRAM集成度高,功耗小,但存取速度慢,一般用来组成大容量主存系统。

ROM是只能随机读出而不能写入的存储器,用来存放那些不需要改变的信息。ROM的结构比RAM简单,集成度高,功耗低,可靠性高。

半导体存储芯片通过地址线、数据线和控制线与外部连接。地址线的数目与芯片容量有关;数据线的数目与芯片数据位数有关;控制线则有读写控制线和片选线等。3种不同类型的半导体存储芯片的对外连接信号有所不同,根据芯片的引脚图就可以区分出这是哪一种类型的芯片,并且能知道此芯片容量的大小和数据的位数。下面以常见的存储芯片为例,讨论各种不同类型存储芯片的特点。

SRAM芯片(如Intel 2114)的引脚为:

地址线—— A_i

数据线—— I/O_i

片选线—— \overline{CS}

读写控制线—— \overline{WE}

电源线: V_{CC} ——+5V,工作电源

GND——地

DRAM芯片(如Intel 2164/4164)的引脚为:

地址线—— A_i

数据线—— D_{IN} 和 D_{OUT}

行地址选通线—— \overline{RAS}

列地址选通线—— \overline{CAS}

读写控制线—— \overline{WE}

电源线: V_{CC} ——+5V,工作电源

GND——地

由于DRAM芯片集成度高、容量大,为了减少芯片引脚数量,DRAM芯片采用了地址复用技术,把地址线分成相等的两部分,分两次从相同的引脚送入。两次输入的地址分别称为行地址和列地址,行地址由行地址选通信号 \overline{RAS} 送入存储芯片,列地址由列地址选通信号 \overline{CAS} 送入存储芯片。因此,DRAM芯片每增加一条地址线,实际上是增加了2位地址,也即增加了4倍的容量。

EPROM 芯片(如 Intel 2732/2764/27128/27256)的引脚:

地址线 — A_i

数据线 — O_i

片选线 — CE

输出允许线 — OE

编程控制线 — PGM

电源线: V_{cc} — +5V, 工作电源

V_{pp} — 编程电源

GND — 地

读方式时, V_{cc} 、 V_{pp} 都接+5V, CE和OE都接低电平(或负脉冲)。

编程方式时, V_{pp} 接+21V, V_{cc} 接+5V, CE为低, OE为高, PGM 端应输入一个宽 50ms 的低电平脉冲, 这是写一个存储单元的时间。

例 5-2 图 5-4 是某存储芯片的引脚图, 请回答:

(1) 这个存储芯片的类型(是 RAM 还是 ROM)?
这个存储芯片的容量?

(2) 若地址线增加一根, 存储芯片的容量将变为多少?

(3) 这个芯片是否需要刷新? 为什么?

解: (1) 从芯片的引脚图可以看出这是一个 DRAM 芯片, 其容量为 $64K \times 1$ 。

(2) 由于 DRAM 采用地址复用技术, 故地址线增加一根, 容量增加 4 倍, 此时存储容量应为 $256K \times 1$ 。

(3) 此芯片需要刷新, 因为 DRAM 是利用栅极电容来存储信息的, 电容上的电荷会随着时间推移而泄漏掉, 必须定时刷新。

8. 存储芯片的地址译码系统

存储芯片的地址译码系统有单译码和双译码两种。

单译码方式仅有一个地址译码器。若译码器的输入线为 n , 则输出线数为 2^n , 即有 2^n 条驱动线(又称字线)对应 2^n 个存储单元。每个存储单元由一条字线选择。

双译码方式有两个译码器: 一个是 X 方向(行)译码器, 一个是 Y 方向(列)译码器。这种方式将存储器地址分成两部分: X 地址(n_x)和 Y 地址(n_y), 即 $n = n_x + n_y$; 若 $n_x = n_y$, 每个译码器的输入为 $k = \frac{n}{2}$ 条线, 则输出为 2^k 条线, 每个译码器有 2^k 条驱动线(驱动器)。每根驱动线驱动一行(列)存储单元, 也就是说每个存储单元由一条 X 方向驱动线和一条 Y 方向驱动线选择。

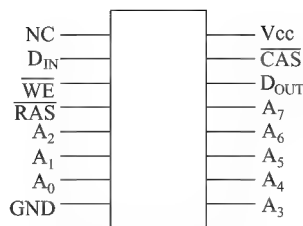


图 5-4 某存储芯片的引脚图

设 $n=16, n_x=n_y=8$ 。对于单译码方式,需要一个 16 位输入的地址译码电路,输出 65 536 条驱动线(驱动器),对应 65 536 个存储单元。对于双译码方式,需要 2 个 8 位输入的地址译码器, X 方向有 256 条驱动线(驱动器), Y 方向也有 256 条驱动线(驱动器),共 512 条驱动线(驱动器)。双译码方式与单译码方式相比,驱动线(驱动器)减少了 99.2%。

9. 存储容量的扩展及存储芯片与 CPU 的连接

主存储器是整个存储系统的核心,通常分为随机存储器(RAM)和只读存储器(ROM)两大部分,RAM 和 ROM 在主存中是统一编址的。有些学生误认为主存中就只有 RAM,这是不正确的。RAM 用来存放供用户随机读写的用户程序和数据,也可以作为系统程序的工作区;ROM 用来存放系统程序。不管是 RAM 还是 ROM,通常都是由许多个不同容量的芯片组成的,其芯片的数量取决于各个存储区域的容量和每个芯片容量的大小。

存储容量的扩展有位扩展、字扩展和字和位同时扩展 3 种。当所用存储芯片中每个单元的位数小于 CPU 访存的字长时,采用位扩展法;当所用存储芯片中每个单元的位数与 CPU 访存的字长相同,但所要求的存储容量大于一片芯片的存储容量时,采用字扩展法;当所用存储芯片的容量和每个单元的位数均不能满足要求时,不仅要位扩展,还需要字扩展,这就是字和位同时扩展。

简单地掌握以上 3 种扩展方式并不是很困难的事,真正的难点在于,如何利用不同类型(RAM 和 ROM)、不同容量的各种存储芯片组成符合特定要求的主存储器。在解决用若干存储芯片构成存储器的问题时,首先要清楚对存储器规模的基本要求,然后要了解可供使用的存储芯片的规格,至此就可以确定这个系统需要多少个芯片,并知道是应当采用位扩展、字扩展还是字和位同时扩展的方法。此时应当把注意力主要放在地址空间的分配和片选逻辑的形成上,这是用存储芯片构成存储器的关键。

对于字扩展、字和位同时扩展的方法必须要考虑各个芯片或各组芯片的地址空间分配问题,建议先依次写出各个芯片或各组芯片在最大存储空间中的地址范围(最低地址和最高地址),接下来再根据地址分配列出芯片的片选逻辑,最后画出连接图。

由存储芯片构成存储器并与 CPU 连接时,要完成:

- 地址线的连接;
- 数据线的连接;
- 控制线的连接。

每个存储芯片上的引脚都是与芯片的类型和容量相对应的,要注意连接到芯片上的地址线、数据线的数量和方向,尤其是当选用的多个芯片容量不同、位数不同时,更要特别地小心。片选逻辑通常会用到译码器或其他的门电路,这要求学生有一定的数字电路的基础。还有一点需要提醒的是,ROM 是只读存储芯片,芯片上没有读写控制引脚,因此

不能将 CPU 的读写控制线接到 ROM 芯片上去。

10. 选片地址的全译码和部分译码

CPU 访问主存时需要首先给出地址码,一般将地址码分成片内地址和选片地址两部分。片内地址由低位的地址码构成,其长度取决于所选存储芯片的字数;选片地址由高位的地址码构成,用于选择存储芯片,大多数情况下由选片地址通过译码后产生存储芯片的片选信号。

选片地址的译码有全译码和部分译码之分。除去片内地址以外的全部地址位都参加译码的方法称为全译码,仅使用高位地址码的部分位来参加译码的方法称为部分译码。

对于实际使用的存储空间小于 CPU 可访问的最大存储空间的情况,全译码意味着实际使用的存储空间的地址范围被严格地限定在最大允许的存储空间中某个一定的区间内,而其他的区域将是空闲的,也就是说,译码器的一些输出端将闲置不接任何存储芯片。全译码方式的最大特点是所使用的存储芯片的地址范围是惟一的,不存在着地址重叠的问题。

对于实际使用的存储空间小于 CPU 可访问的最大存储空间的情况,部分译码则仅使用高位地址码的部分位来译码,剩余的地址位不参与译码,此时存储芯片的地址范围不再是惟一的,而出现了重叠的地址。其地址重叠区的个数取决于没有参加译码的高位地址码的位数,如果有 2 位地址没有参加译码,则就会出现 4 个地址重叠区;有 3 位地址没有参加译码,则就会出现 8 个地址重叠区;……。

5.3 相关知识介绍

1. 存储系统的性能分析

表示存储系统的性能有 3 个主要参数:容量 S , 价格 C 和速度 T , 组成这个存储系统的每个存储器也有同样的 3 个参数。图 5-5 表示由两个存储器 M_1 和 M_2 组成的存储系统,两个存储器的容量、价格和速度分别为 S_1 、 C_1 、 T_1 和 S_2 、 C_2 、 T_2 。

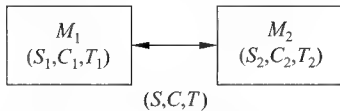


图 5-5 由两个存储器构成的存储系统

整个存储系统的单位容量平均价格可以表示为:

$$C = \frac{C_1 S_1 + C_2 S_2}{S_1 + S_2}$$

当 $S_2 \gg S_1$ 时, 则有 $C \approx C_2$, 因此, 整个存储系统的单位容量价格接近于比较便宜的

M_2 存储器。

存储系统的访问周期与命中率 H 的关系比较大,命中率可以简单地定义为在 M_1 存储器中访问到的概率。即:

$$H = \frac{N_1}{N_1 + N_2}$$

其中, N_1 表示对 M_1 存储器的访问次数, N_2 表示对 M_2 存储器的访问次数。

整个存储系统的访问周期可以用 M_1 和 M_2 两个存储器的访问周期 T_1 、 T_2 和命中率 H 来表示:

$$T = H \times T_1 + (1 - H) \times T_2$$

当命中率 $H \rightarrow 1$ 时, $T \rightarrow T_1$, 即存储系统的访问周期 T 接近于速度比较快的 M_1 存储器的访问周期。

为了便于分析,定义存储系统的访问效率为:

$$e = \frac{T_1}{H \times T_1 + (1 - H) \times T_2} = \frac{1}{H + (1 - H) \times \frac{T_2}{T_1}} = f\left(H, \frac{T_2}{T_1}\right)$$

访问效率越高,说明存储系统的速度与相对比较快的那个存储器速度越接近。从上述式式可以看到,存储系统的访问效率主要与命中率和构成存储系统的两级存储器的速度之比有关。要使存储系统的速度接近相对比较快的那个存储器速度的途径有:①提高命中率,②使构成存储系统的两个存储器的速度不要相差太大。通常 Cache 存储系统中两个存储器的速度之比 $\frac{T_2}{T_1}$ 比较小,约为 3~10 倍,而虚拟存储系统中两个存储器的速度之比 $\frac{T_2}{T_1}$ 比较大,约为 10^5 倍,所以说,虚拟存储系统如果要想获得比较高的访问效率,则需要极高的命中率。

2. 访问的局部性原理

程序往往重复使用它刚刚使用过的数据和指令。实验表明,一个程序用 90% 的执行时间去执行仅占 10% 的程序代码。局部性分为时间上的局部性和空间上的局部性两种。时间上的局部性是指如果一个存储单元被访问,则可能该单元会很快被再次访问,这是因为程序存在着循环。空间上的局部性是指如果一个存储单元被访问,则该单元邻近的单元也可能很快被访问。这是因为程序中大部分指令是顺序存储、顺序执行的,数据一般也是以向量、数组、树、表等形式簇聚地存储在一起的。也就是说,最近的、未来要用的指令和数据大多局限于正在用的指令和数据,或是存放在与这些指令和数据位置上邻近的单元中。这样,就可以把目前常用或将要用到的信息预先放在 M_1 中,从而使 CPU 的访问速度大大提高。存储系统的构成就是以访问的局部性原理为基础的。

3. 静态 RAM 芯片分析

Intel 2114 是一种曾经广泛使用的小容量 SRAM 芯片,它的结构框图如图 5-6 所示,由存储矩阵、行/列地址选择电路、列 I/O 电路及三态读写电路组成。

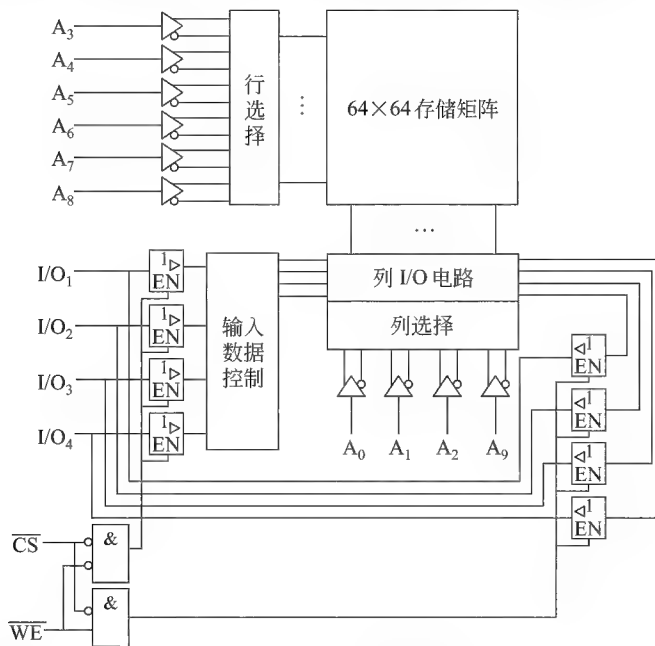


图 5-6 2114 逻辑结构框图

2114 的存储容量为 $1K \times 4$, 由 4096 个 6 管记忆单元电路组成, 它们排成 64×64 的矩阵, 采用字段结构。64 列被分成 4 组, 每组包含 16 列。第一组的 $64 \text{ 行} \times 16 \text{ 列}$ 中的各记忆单元表示 1024 个存储单元中的第一位, 第二组的 $64 \text{ 行} \times 16 \text{ 列}$ 中的各记忆单元表示 1024 个存储单元中的第二位, …… , 依次类推, 便构成了存储芯片的 4 位, 2114 芯片存储体内部结构如图 5-7 所示。

CPU 送来的地址总线共 10 位 ($A_0 \sim A_9$), 其中 6 位 ($A_3 \sim A_8$) 作为行选择电路的输入, 经行地址译码器产生 64 条行选择线 ($X_0 \sim X_{63}$); 另 4 位 ($A_0 \sim A_2, A_9$) 作为列选择电路的输入, 经列地址译码器产生 16 条列选择线 ($Y_0 \sim Y_{15}$)。当任何一条 X 选择线和 Y 选择线被选时, 其交点处的 4 个记忆单元被选。

$I/O_1 \sim I/O_4$ 是受输入三态门和输出三态门控制的双向数据线, 由片选信号 \overline{CS} 和写允许信号 \overline{WE} 一起来控制这些三态门。在 \overline{CS} 有效 (低电平) 的情况下, 如果 \overline{WE} 有效 (低电平), 则输入三态门打开, 数据总线上的信息便写入存储器; 如果 \overline{WE} 无效, 则打开输出三态门, 信息从存储器中读出来, 送到数据总线上。由于读或写操作是分时进行的, 即读时

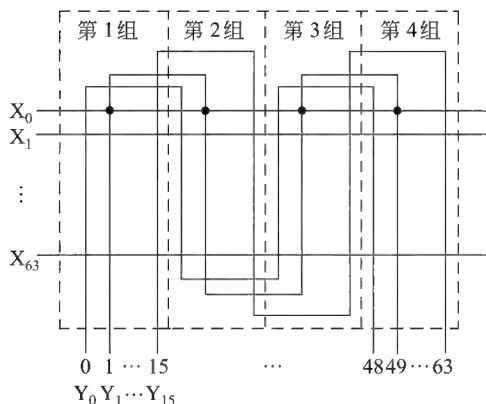


图 5 7 2114 芯片存储体内部结构

不写,写时不读,因此,输入和输出三态门是互锁的。

4. 动态 RAM 芯片分析

Intel 2164/4164 是 $64\text{K} \times 1$ 的 DRAM 芯片,片内集成有 65 536 个单管 MOS 动态记忆单元,2164/4164 芯片的逻辑结构框图如图 5-8 所示。

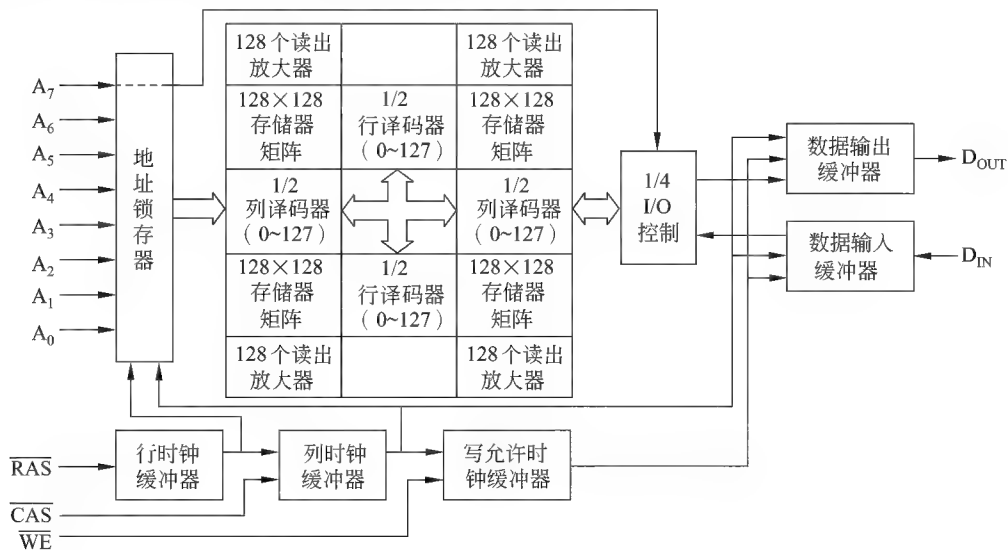


图 5 8 2164/4164 逻辑结构框图

2164/4164 芯片的存储体分成 4 个 128×128 的存储矩阵。选择 64K 容量本来需要 16 位地址,而 4164 芯片的地址线引脚只有 8 条 ($A_7 \sim A_0$),需分时复用。4164 内部设有 8

位的地址锁存器,将 16 位的地址信息分成两次送到 4164 内,共同译码指定被访问的存储单元。低 8 位地址作为行地址,高 8 位地址作为列地址,行/列地址转换控制电路如图 5-9 所示。图中,ADDR SEL 是行/列地址转换控制信号,当它为 0 时,地址码的低 8 位 $XA_{7\sim0}$ 通过多路选择器;为 1 时,地址码的高 8 位 $XA_{15\sim8}$ 通过多路选择器。

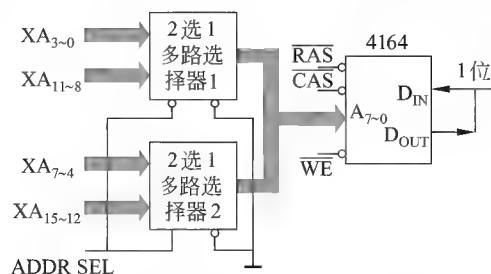


图 5-9 行/列地址转换控制电路

行地址由行地址选通信号 RAS 送至地址锁存器,其中低 7 位经过译码后产生 128 条行选择线,可选择 128 行中的任一行。接着,列地址由列选通信号 CAS 送至地址锁存器,其中低 7 位列地址经过译码后产生 128 条列选择线,分别选择 128 列中的任一列。这时 4 个存储矩阵中各有一位与 I/O 控制电路连接,最后将行地址和列地址的最高位 (A_7 和 A_{15}) 送到 I/O 控制电路,用以选择 4 个矩阵中的一个矩阵。当 \overline{WE} 为高电平时,把 16 位地址所指定的单元中的数据通过数据输出缓冲器送到 D_{OUT} 端,当 \overline{WE} 为低电平时, D_{IN} 端的数据通过数据输入缓冲器输入,写入到指定的单元中。

4164 中每一列都有读出放大器,这是因为芯片的记忆单元是由单个 MOS 管和电容组成的,其读出信号是很微弱的。

4164 的刷新操作是通过执行只有 RAS 的存取周期来实现的,这时 4164 只取 8 位地址码中的低 7 位 $A_6\sim A_0$,4 个 128×128 的存储矩阵中凡是 $A_6\sim A_0$ 为符合给定地址码的所有 512 个单元全部刷新。即 16 位地址码中, $A_6\sim A_0$ 为给定地址码, $A_{15}\sim A_7$ 为任意项,刷新过程如下:

A_6	A_5	A_4	A_3	A_2	A_1	A_0	刷新的 512 个单元
0	0	0	0	0	0	0	0000H,0080H,0100H,0180H,...,FF00H,FF80H
0	0	0	0	0	0	1	0001H,0081H,0101H,0181H,...,FF01H,FF81H
0	0	0	0	0	1	0	0002H,0082H,0102H,0182H,...,FF02H,FF82H
...							
1	1	1	1	1	1	1	007FH,00FFH,017FH,01FFH,...,FF7FH,FFFFH

$A_6\sim A_0$ 可由 DMA 控制器提供,当 $A_6\sim A_0$ 从 0000000 变化到 1111111 时,4164 的内容就刷新了一遍。4164 刷新一遍所需时间为 2ms,在 2ms 内执行 128 次刷新操作,要求

每次操作的时间间隔为 $15.6\mu\text{s}$ 。由于刷新时没用 CAS 信号,故与外界不发生数据传送。

5. MROM 和 PROM 的写入

ROM 在一般情况下的只能读出,不能写入,那么 ROM 中的内容是如何事先存入的呢? 下面分析 MROM 和 PROM 的写入原理。

(1) 掩膜式 ROM (MROM)

掩膜式 ROM 的内容是由半导体制造厂按用户提出的要求在芯片的生产过程中直接写入的,写入之后任何人都无法改变其内容。MROM 中的记忆单元可采用二极管、电阻、双极型晶体管、MOS 管等作为耦合元件。通常凡耦合处有元件表示存储“0”信息,无元件表示存储“1”信息。

图 5-10 为一个简单的 4×4 位 MOS 管 ROM,采用单译码结构,两位地址线 A_1 、 A_0 译码后可有 4 种状态,驱动 4 条选择线,可分别选中 4 个单元,每个单元有 4 位输出。

图 5-10 所示的矩阵中,在行和列的交点处,既可有耦合元件 MOS 管,也可没有。若地址线 $A_1 A_0 = 00$,则选中 0 号单元,即字线 0 为高电平,其他字线为低电平。如果位线上有 MOS 管与字线 0 相连(如位线 2 和 0),其对应的 MOS 管导通,位线输出为“0”;如果位线上没有 MOS 管与字线相连(如位线 1 和 3),则位线输出为“1”。

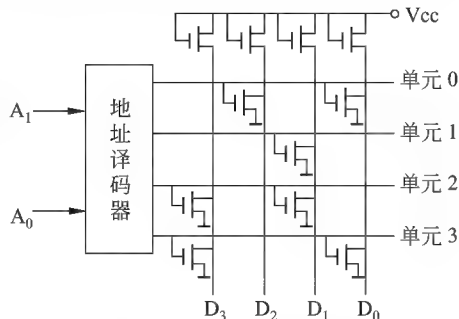
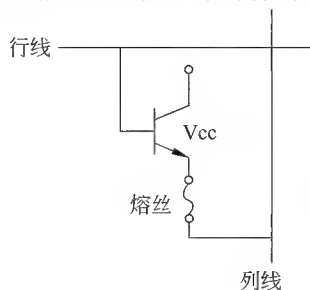


图 5-10 4×4 位掩膜式 ROM

(2) 一次可编程 ROM (PROM)

PROM 产品出厂时,所有记忆单元均制成“0”(或制成“1”),用户根据需要可自行将其中某些记忆单元改为“1”(或改为“0”)。常见的 PROM 根据写入原理可分为两类:结破坏型和熔丝型。由于它们的写入都是不可逆的,所以只能进行一次性写入。

结破坏型在每个行、列线的交点处,制造一对彼此反向的二极管,它们因为彼此反向而不能导通,故全部内容均为“0”。若某位需要写入“1”,则在相应的行、列之间加上较高电压,将反偏的一只二极管永久性击穿,只留下正向导通的一只二极管,故该位被写入“1”。显然这种写入是一次性的,不可逆转的。



熔丝型的基本记忆单元电路是由三极管 T 连接一段镍—铬熔丝组成的,见图 5-11。典型的 PROM 芯片出厂时,T 与位线之间的熔丝都存在,表示全部内容均为“0”。当用户需要某一位写入“1”时,则设法将 T 管的电流加大为正常工作电流的五倍以上,从而使镍—铬熔丝熔断,“1”

被写入。由于熔丝熔断之后不能再恢复,显然这种写入也是不可逆转的。

6. RAM 的奇偶校验电路

为了检测存储过程中的错误, RAM 中最常用的是奇偶校验方法。例如,用 4164 芯片组成的 64KB 存储器的奇偶校验电路如图 5-12 所示。

从图 5-12 中可以看出,该存储器由 9 片 4164 组成,其中 1~8 片组成 $64K \times 8$ 存储器,第 9 片用来作奇偶校验。该电路的核心是 74LS280,它有 9 个输入端(A~I)和两个互非的输出端(EVEN 和 ODD)。当输入端 1 的个数为偶数时,EVEN 为高电平,ODD 为低电平;当输入端 1 的个数为奇数时,ODD 为高电平,EVEN 为低电平。

奇偶校验(以奇校验为例)的原理是:写操作时,存储器读信号 $XMEMR = 1$,使 LS280 的 I 输入端为 0,这样当 8 位数据中 1 的个数为偶数时,便使第 9 片 4164 的相应单元写入 1,否则写入 0。当读这个单元的数据时, $XMEMR = 0$,若所存的 8 位数据没有发生读出错误,就使 $ODD = 1, U_{ODD} = 0$;若发生读出错误,则使 $ODD = 0, U_{ODD} = 1$ 。也就是说, U_{ODD} 的输出即可以判断有无奇偶错误,此信号向 CPU 发出奇偶校验错的中断请求信号。

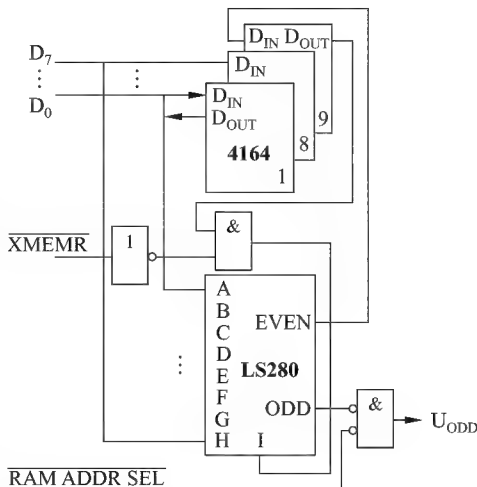


图 5 12 RAM 的奇偶校验电路

7. BIOS 和 CMOS 芯片

BIOS 是“Basic Input/Output System”的缩写,意思是“基本输入输出系统”,通常固化在只读存储器(ROM)中,所以又称为 ROM-BIOS。ROM-BIOS 是计算机系统中用来提供最低级、最直接的硬件控制的程序,是连接软件程序和硬件设备之间的枢纽,可以直接对计算机系统中的输入输出设备进行设备级、硬件级的控制。当计算机开机时,首先执行的是 BIOS 中的程序,它的功能是上电自检、开机引导、基本外设 I/O 和系统的 CMOS 设置。

目前 BIOS 的容量越来越大,其中的新技术也是越来越多,如 BIOS 里面固化了防病毒技术、双 BIOS 芯片技术,以备份的形式防止病毒的侵袭、在线更新主板 BIOS 的功能等。

CMOS 是“Complementary Metal Oxide Semiconductor”的缩写,其本意是“互补金属氧化物半导体”,这是一种应用于集成电路芯片制造的原料。但我们在接触主板时说的这个“CMOS”则是指主板上一种低功耗的、靠电池供电的可读写存储器(RAM)芯片,用来

保存当前系统的硬件配置和用户对某些参数的设定。CMOS 耗电量非常低,即使系统断电或者长期不开机,CMOS 中的信息也不会丢失。

BIOS 与 CMOS 既相关又不同。BIOS 中的系统设置程序是完成参数设置的手段,CMOS 是设定系统参数的存放场所。它们跟系统设置都密切相关,正确的说法是:当进入 BIOS 对硬盘参数或者其他 BIOS 进行设置并保存它们时,这些设置会被存储到 CMOS RAM 芯片中去。每次系统引导时,系统都会从 CMOS RAM 芯片中读出所存的参数来决定如何配置系统,BIOS 和 CMOS RAM 之间存在联系,但它们是系统中两个完全不同的部分。

8. 双端口存储器

常规存储器是单端口存储器,每次只接收一个地址,访问一个存储单元,从中读取或存入一个字节或一个字。在执行双操作数指令时,就需要分两次读取操作数,工作速度较低。在高速系统中,主存储器是信息交换的中心,一方面 CPU 频繁地与主存交换信息,从中读取指令、存取数据,另一方面外设也需较频繁地与主存交换信息。而单端口存储器每次只能接受一个访存者,或是读或是写,这就影响了工作速度。为此,在某些系统或部件中使用双端口存储器。

图 5-13 所示双端口存储器具有两个彼此独立的读/写口,每个读/写口都有一套独立的地址寄存器和译码电路,可以并行地独立工作。两个读/写口可以按各自接收的地址,同时读出或写入,或一个写入而另一个读出。与两个独立的存储器不同,两套读/写口的访存空间相同,可以访问同一区间、同一存储单元。

双端口存储器的常见应用场合有:在运算器中采用双端口存储芯片,作为通用寄存器组,能够快速提供双操作数,或快速实现寄存器间传送。另一种应用是让双端口存储器的一个读/写口面向 CPU,通过专门的存储总线(或称局部总线)连接 CPU 与主存,使 CPU 快速访问主存;另一个读/写口则面向外设或输入输出处理机,通过共享的系统总线连接。如显示器上使用的视频 DRAM(又称 VRAM),它的两个端口中,一个允许 CPU 随机读写,另一个只能被视频显示电路读取,且一次只能读一整行。此外,在多机系统中常用双端口存储器甚至多端口存储器,作为各 CPU 的共享存储器,实现多 CPU 之间的通信。

9. 多体并行系统

n 个并行的存储器具有各自的地址寄存器(MAR)、读/写电路和数据寄存器(MDR),它们能各自以同等的方式与 CPU 传递信息,形成可以同时工作又独立编址且容量相同的 n 个分存储体,这就是多体系统。各个存储体能并行工作,也能分时交叉工作。

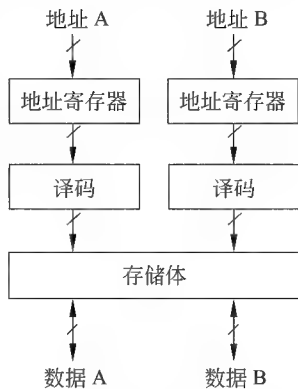


图 5-13 双端口存储器

并行工作即同时访问 n 个存储体,同时启动,同时读出,完全并行地工作;分时工作即 n 个存储体以 $\frac{T_m}{n}$ 的时间间隔进入并行工作状态。

图 5-14 所示的是高位交叉编址的多体存储系统。存储器地址寄存器的高位表示体号,低位表示体内地址。程序按体内地址连续存放,一个体存满后,再存入下一个体。

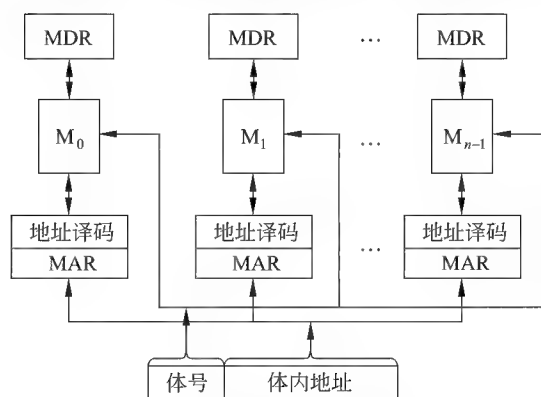


图 5 14 高位交叉编址的多体存储系统

高位交叉编址时,系统地址的连续地址落在同一存储体内,容易发生访存冲突,并行存取的可能性很小。

图 5-15 是按低位交叉编址的多体存储系统。同一存储体中的地址是不连续的,程序连续存放在相邻体中。存储器地址寄存器的低位部分选择不同的存储体,而高位部分则指向存储体内的存储字。

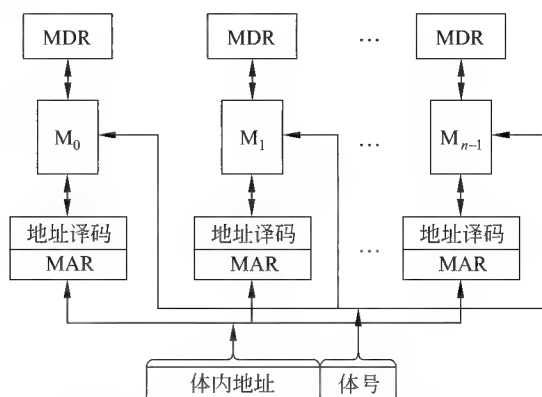


图 5 15 低位交叉编址的多体存储系统

低位交叉编址时,系统地址在同一存储体中不是连续的,而是以 n 为模交叉编址的。所以连续的程序或数据将交叉的存放在 n 个存储体中,可实现以 n 为模的交叉并行存取,访存冲突的概率就会变得很小。

目前大多数计算机都采用多体低位交叉编址并行主存,这将大大提高主存频宽,从而提高计算机系统的性能。并行主存的实际频宽总是小于最大频宽,换句话说,访存冲突总是存在的。究其原因,除了程序不总是顺序执行和数据随机存放之外,还与存储体个数一般为 2 的整数次幂有关。可以证明当存储体个数 n 取 5 以上的素数时,访存冲突将大大减小,且 n 取的素数越大,访存冲突越少。因此称采用素数个存储体的低位交叉并行主存为无访问冲突并行主存,其实际频宽接近于最大频宽。无冲突并行主存,因存储体个数为素数,由系统地址变成体号和体内地址是很复杂的。

10. 主存与 Cache 之间的地址变换与映像方式

在 Cache 中,地址映像是指把主存地址空间映像到 Cache 地址空间,具体地说,就是把存放在主存中的程序按照某种规则装入到 Cache 中,并建立主存地址与 Cache 地址之间的对应关系。而地址变换则是指当程序已经装入到 Cache 之后,在实际运行过程中,主存地址如何变换成 Cache 地址。

无论采用什么样的地址映像和地址变换方式,都要把主存和 Cache 划分成同样大小的块,每块可包含几十个或几百个存储字,显然主存中的块数会比 Cache 中的块数多得多。例如,某系统中主存容量为 1MB,而 Cache 容量为 8KB,每 1KB 为一块,于是主存中共有 1024 块,而 Cache 中只有 8 块,这就是说,任何时候主存中最多只能有 8 块信息进入 Cache 中。

地址映像的方式有 3 种:全相联映像、直接映像和组相联映像。全相联映像方式是指主存的任何一块可调入 Cache 的任何一块位置,主存和 Cache 之间的地址变换简化成为块号的变换。上例中主存有 1024 块,Cache 共 8 块,主存和 Cache 的地址结构如图 5-16(a)所示。直接映像方式是指主存中的某些块只能固定的调入 Cache 中的某一块位置,它们之间存在着固定的关系。上例中 Cache 共 8 块,而主存有 1024 块,可分成 128 个区,主存和 Cache 之间的地址结构如图 5-16(b)所示。组相联映像方式将主存和 Cache 分成同样大小的组,每个组内包含同样数量的块,组内采用全相联映像,组间采用直接映像方式。上例中 Cache 每 4 块构成一个小组,共两个小组;主存中同样每 4 块构成一个小组,两个小组构成一个区,于是主存中共有 128 个区。主存和 Cache 的地址结构如图 5-16(c)所示。

11. 虚拟存储器的工作过程

在虚拟存储器中有 3 个地址空间,一是虚拟地址空间,也称虚存空间和虚拟存储器空间,它是应用程序员用来编写程序的地址空间,这个地址空间非常大;二是主存储器的地址空间,也称主存地址空间、主存物理空间或实存地址空间;三是辅存地址空间,也就是磁

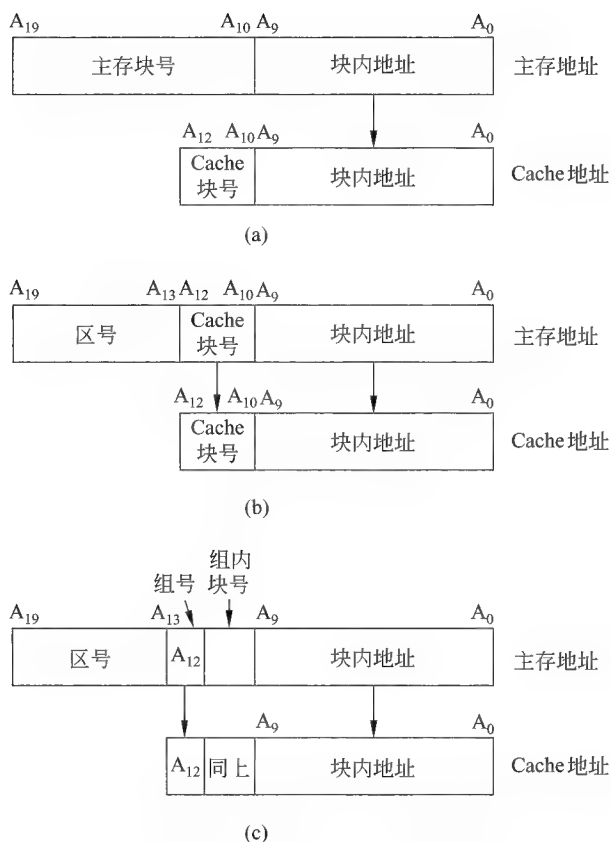


图 5 16 主存和 Cache 的地址结构

盘存储器的地址空间。与这 3 个地址空间相对应,有 3 种地址,即虚拟地址(虚存地址、虚地址)、主存地址(主存实地址、主存物理地址、主存储器地址)和磁盘存储器地址(磁盘地址、辅存地址)。

地址映像是把虚拟地址空间映像到主存地址空间,具体地说,就是把用户用虚拟地址编写的程序按照某种规则装入到主存储器中,并建立多用户虚地址与主存实地址之间的对应关系。而地址变换则是在程序装入主存储器之后,在实际运行时,把多用户虚地址变换成主存实地址(内部地址变换)或磁盘存储器地址(外部地址变换)。

下面以页式虚拟存储器为例,讨论虚拟存储器的工作过程,其工作原理如图 5-17 所示。

在页式虚拟存储器的访问过程中,可能会用到 3 张表。一是页表,也称为内页表,它是在内部地址变换中使用的,每个用户都有一张内页表,表中应包含虚页号、实页号、装入位等信息,装入位为“1”,表示该实页已被占用;二是外页表,它是在外部地址变换中使用的,每个用户都有一张外页表,其内容至少应包括辅存实页号(磁盘地址)和装入位,如果装入位为“1”,表示要访问的页面已经在磁盘存储器中,否则表示不在磁盘存储器中,需要从其他海量

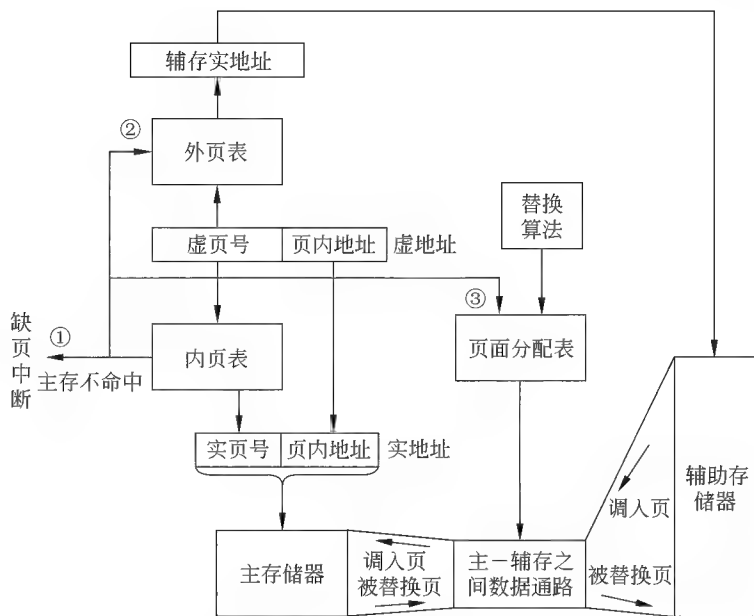


图 5-17 页式虚拟存储器的工作过程

存储器中调入;三是整个系统还需要设置一个页面分配表,供所有用户公用,这个表用来记录当前各个实存页面的使用情况,表中至少要包含用户号、装入位、历史位、修改位等信息。

虚拟存储器工作的过程是:CPU 用虚地址访问存储器,首先查内页表,判断实页是否命中。若命中,则从内页表中得到实页号,并将其与页内地址拼接起来构成访问实存的实地址;若不命中,则需要完成 3 项任务:①向 CPU 发出缺页中断;②查外页表得到该页的辅存实页号,并用该页号去访问辅存,将该页调入实存中去,并填写好外页表;③查实存页面分配表,若实存中还有空闲页面,则将从辅存中取出的页面直接写入实存的空闲页面中,并填写好内页表,若实存空间已满,则需根据所采用的替换算法确定当前的被替换页面。若该页面内容在执行中被修改过,则需先将它写回到辅存中原来所在的页面,然后才能将新页调入实存覆盖掉被替换页,若未曾修改过,则可将新页直接覆盖被替换页,并填写好内页表,然后才能用原虚地址去访问实存,这时实存肯定命中,经内部地址变换后,可直接去访问实存,完成一次访问虚拟存储器的全过程。

12. 相联存储器

常规存储器是按地址访问的,即送一个地址码,选中相应的一个编址单元,然后进行读写操作。而在信息检索一类工作中,需要的却是按信息内容选中相应单元,进行读写。相联存储器又称为联想存储器,它不是根据地址访问存储器,而根据所存数据字的全部内容或部分内容进行存取的,是一种按内容寻址的存储器。相联存储器的基本组成如图 5-18 所示。设存储器有 W 个字,字长 n 位,CR 为比较数寄存器,字长也为 n 位,存放要比

较的数(或要检索的内容)。MR 为屏蔽寄存器,与 CR 配合使用,字长也为 n 位。当按比较数的部分内容进行检索时,相应地把 MR 中要比较的位设置成 1,不要比较的位设置成 0。图 5-18 中表示需要按 2~6 位的内容进行比较,所以 MR 的第 2~6 位置 1,其余各位均置 0。置成 1 的字段称为关键字段。SRR 为查找结果寄存器,字长为 W 位,假如比较结果第 i 个字满足要求,则 SRR 中的第 i 位为 1,其余各位均为 0,若同时有 m 个字满足要求,则相应地就有 m 位为 1。有的相联存储器还设置有字选择寄存器 WSR,用来确定哪些字参与检索,若字选择寄存器某位为 1,则表示其对应的存储字参与检索;若某位为 0,则表示其对应的存储字不参与检索。

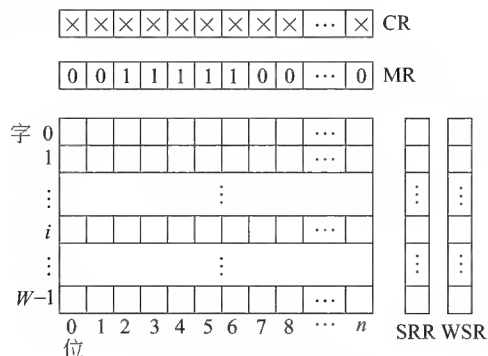


图 5 18 相联存储器框图

为了进行检索,还要求相联存储器能进行各种比较操作(相等、不等、小于、大于、求最大值和最小值等)。比较操作是并行进行的,即 CR 值的关键字段与存储器的所有 W 个字的相应字段同时进行比较。

相联存储器常用于虚拟存储器和 Cache 中,还经常用于数据库与知识库中按关键字进行检索。从按地址访问的存储器中检索某一个单元,平均约进行 $W/2$ 次操作(W 为存储单元数),而在相联存储器中仅需要进行一次检索操作,所以大大提高了处理速度。

5.4 教材习题解答

1. 如何区别存储器和寄存器? 两者是一回事的说法对吗?

解: 存储器和寄存器不是一回事。存储器在 CPU 的外边,专门用来存放程序和数据,访问存储器的速度较慢。寄存器属于 CPU 的一部分,访问寄存器的速度很快。

2. 存储器的主要功能是什么? 为什么要把存储系统分成若干个不同层次? 主要有哪些层次?

解: 存储器的主要功能是用来保存程序和数据。存储系统是由几个容量、速度和价

格各不相同的存储器用硬件、软件、硬件与软件相结合的方法连接起来的系统。把存储系统分成若干个不同层次的目的是为了解决存储容量、存取速度和价格之间的矛盾。由高速缓冲存储器、主存储器、辅助存储器构成的一级存储系统可以分为两个层次,其中高速缓存和主存间称为 Cache—主存存储层次(Cache 存储系统);主存和辅存间称为主存—辅存存储层次(虚拟存储系统)。

3. 什么是半导体存储器? 它有什么特点?

解: 采用半导体器件制造的存储器,主要有 MOS 型存储器和双极型存储器两大类。半导体存储器具有容量大、速度快、体积小、可靠性高等特点。半导体随机存储器存储的信息会因为断电而丢失。

4. SRAM 记忆单元电路的工作原理是什么? 它和 DRAM 记忆单元电路相比有何异同点?

解: SRAM 记忆单元由 6 个 MOS 管组成,利用双稳态触发器来存储信息,可以对其进行读或写,只要电源不断电,信息将可保留。DRAM 记忆单元可以由 4 个和单个 MOS 管组成,利用栅极电容存储信息,需要定时刷新。

5. 动态 RAM 为什么要刷新? 一般有几几种刷新方式? 各有什么优缺点?

解: DRAM 记忆单元是通过栅极电容上存储的电荷来暂存信息的,由于电容上的电荷会随着时间的推移被逐渐泄放掉,因此每隔一定的时间必须向栅极电容补充一次电荷,这个过程就叫做刷新。

常见的刷新方式有集中式、分散式和异步式 3 种。集中方式的特点是读写操作时不受刷新工作的影响,系统的存取速度比较高;但有死区,而且存储容量越大,死区就越长。分散方式的特点是没有死区;但它加长了系统的存取周期,降低了整机的速度,且刷新过于频繁,没有充分利用所允许的最大刷新闻隔。异步方式虽然也有死区,但比集中方式的死区小得多,而且减少了刷新次数,是比较实用的一种刷新方式。

6. 一般存储芯片都设有片选端 \overline{CS} , 它有什么用途?

解: 片选线 \overline{CS} 用来决定该芯片是否被选中。 $\overline{CS}=0$, 芯片被选中; $\overline{CS}=1$, 芯片不选中。

7. DRAM 芯片和 SRAM 芯片通常有何不同?

解: 主要区别有:

① DRAM 记忆单元是利用栅极电容存储信息;SRAM 记忆单元利用双稳态触发器来存储信息。

② DRAM 集成度高,功耗小,但存取速度慢,一般用来组成大容量主存系统;SRAM 的存取速度快,但集成度低,功耗也较大,所以一般用来组成高速缓冲存储器和小容量主存系统。

③ SRAM 芯片需要有片选端 \overline{CS} , DRAM 芯片可以不设 \overline{CS} , 而用行选通信号 \overline{RAS} 、列

选通CAS兼作片选信号。

④ SRAM 芯片的地址线直接与容量相关,而 DRAM 芯片常采用了地址复用技术,以减少地址线的数量。

8. 有哪一种只读存储器? 它们各自有何特点?

解: MROM:可靠性高,集成度高,形成批量之后价格便宜,但用户对制造厂的依赖性过大,灵活性差。

PROM:允许用户利用专门的设备(编程器)写入自己的程序,但一旦写入后,其内容将无法改变。写入都是不可逆的,所以只能进行一次性写入。

EPROM:不仅可以由用户利用编程器写入信息,而且可以对其内容进行多次改写。EPROM 又可分为两种:紫外线擦除(UVEPROM)和电擦除(EEPROM)。

闪存存储器:既可在不加电的情况下长期保存信息,又能在线进行快速擦除与重写,兼备了 EEPROM 和 RAM 的优点。

9. 说明存取周期和存取时间的区别。

解: 存取周期是指主存进行一次完整的读写操作所需的全部时间,即连续两次访问存储器操作之间所需要的最短时间。存取时间是指从启动一次存储器操作到完成该操作所经历的时间。存取周期一定大于存取时间。

10. 一个 $1K \times 8$ 的存储芯片需要多少根地址线、数据输入线和输出线?

解: 需要 10 根地址线,8 根数据输入和输出线。

11. 某机字长为 32 位,其存储容量是 64KB,按字编址的寻址范围是多少? 若主存以字节编址,试画出主存字地址和字节地址的分配情况。

解: 某机字长为 32 位,其存储容量是 64KB,按字编址的寻址范围是 16KW。若主存以字节编址,每一个存储字包含 4 个单独编址的存储字节。假设采用大端方案,即字地址等于最高有效字节地址,且字地址总是等于 4 的整数倍,正好用地址码的最末两位来区分同一个字中的 4 个字节。主存字地址和字节地址的分配情况如图 5-19 所示。

字地址	字节地址			
	0	1	2	3
0	0	1	2	3
4	4	5	6	7
8	8	9	10	11
⋮	⋮	⋮	⋮	⋮
65532	65532	65533	65534	65535

图 5-19 主存字地址和字节地址的分配

12. 一个容量为 $16K \times 32$ 位的存储器,其地址线和数据线的总和是多少? 当选用下列不同规格的存储芯片时,各需要多少片?

1K×4 位, 2K×8 位, 4K×4 位, 16K×1 位, 4K×8 位, 8K×8 位。

解: 地址线 14 根, 数据线 32 根, 共 46 根。

若选用不同规格的存储芯片, 则需要: 1K×4 位芯片 128 片, 2K×8 位芯片 32 片, 4K×4 位芯片 32 片, 16K×1 位芯片 32 片, 4K×8 位芯片 16 片, 8K×8 位芯片 8 片。

13. 现有 1024×1 的存储芯片, 若用它组成容量为 16K×8 的存储器。试求:

(1) 实现该存储器所需的芯片数量?

(2) 若将这些芯片分装在若干块板上, 每块板的容量为 4K×8, 该存储器所需的地址线总位数是多少? 其中几位用于选板? 几位用于选片? 几位用作片内地址?

解: (1) 需 1024×1 的芯片 128 片。

(2) 该存储器所需的地址线总位数是 14 位, 其中 2 位用于选板, 2 位用于选片, 10 位用作片内地址。

14. 已知某机字长 8 位, 现采用半导体存储器作主存, 其地址线为 16 位, 若使用 1K×4 的 SRAM 芯片组成该机所允许的最大主存空间, 并采用存储模板结构形式。

(1) 若每块模板容量为 4K×8, 共需多少块存储模板?

(2) 画出一个模板内各芯片的连接逻辑图。

解: (1) 根据题意可知存储器容量为 $2^{16} = 64\text{KB}$, 故共需 16 块存储模板。

(2) 一个模板内各芯片的连接逻辑图如图 5-20 所示。

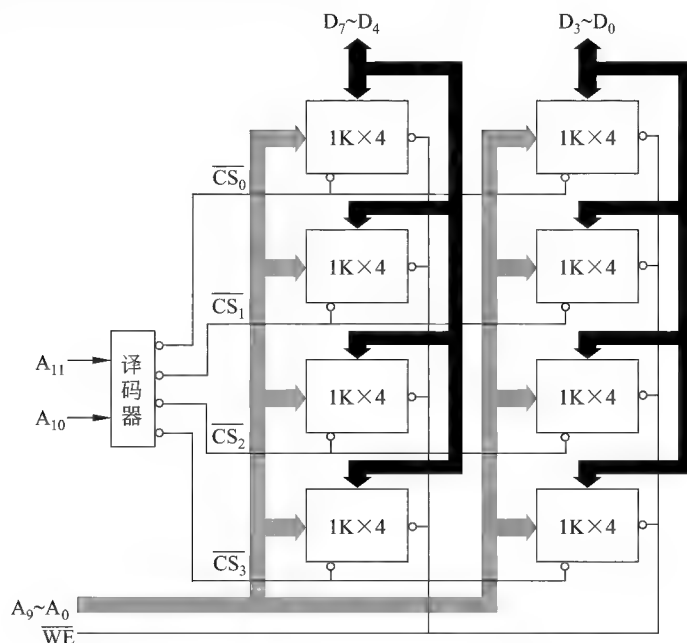


图 5 20 模板内各芯片的连接逻辑图

15. 某半导体存储器容量 $16K \times 8$, 可选 SRAM 芯片的容量为 $4K \times 4$; 地址总线 $A_{15} \sim A_0$ (低), 双向数据总线 $D_7 \sim D_0$ (低), 由 R/W 线控制读/写。请设计并画出该存储器的逻辑图, 并注明地址分配、片选逻辑及片选信号的极性。

解: 存储器的逻辑图与图 5-20 很相似, 区别仅在于地址线的连接上, 故省略。

地址分配如下:

A_{15}	A_{14}	A_{13}	A_{12}	$A_{11} \sim A_0$	
X	X	0	0	——	第一组
X	X	0	1	——	第二组
X	X	1	0	——	第三组
X	X	1	1	——	第四组

假设采用部分译码方式, 片选逻辑为:

$$\overline{CS_0} = \overline{A_{13}} \cdot \overline{A_{12}}$$

$$\overline{CS_1} = \overline{A_{13}} \cdot A_{12}$$

$$\overline{CS_2} = A_{13} \cdot \overline{A_{12}}$$

$$\overline{CS_3} = A_{13} \cdot A_{12}$$

16. 现有如下存储芯片: $2K \times 1$ 的 ROM、 $4K \times 1$ 的 RAM、 $8K \times 1$ 的 ROM。若用它们组成容量为 16KB 的存储器, 前 4KB 为 ROM, 后 12KB 为 RAM, CPU 的地址总线 16 位。

(1) 各种存储芯片分别用多少片?

(2) 正确选用译码器及门电路, 并画出相应的逻辑结构图。

(3) 指出有无地址重叠现象。

解: (1) 需要用 $2K \times 1$ 的 ROM 芯片 16 片, $4K \times 1$ 的 RAM 芯片 24 片。

不能使用 $8K \times 1$ 的 ROM 芯片, 因为它大于 ROM 应有的空间。

(2) 各存储芯片的地址分配如下:

A_{15}	A_{14}	A_{13}	A_{12}	A_{11}	$A_{10} \sim A_0$	
X	X	0	0	0	——	2KB ROM
X	X	0	0	1	——	2KB ROM
X	X	0	1	——	——	4KB RAM
X	X	1	0	——	——	4KB RAM
X	X	1	1	——	——	4KB RAM

相应的逻辑结构图如图 5-21 所示。

(3) 有地址重叠现象。因为地址线 A_{15} 、 A_{14} 没有参加译码。

17. 用容量为 $16K \times 1$ 的 DRAM 芯片构成 64KB 的存储器。

(1) 画出该存储器的结构框图。

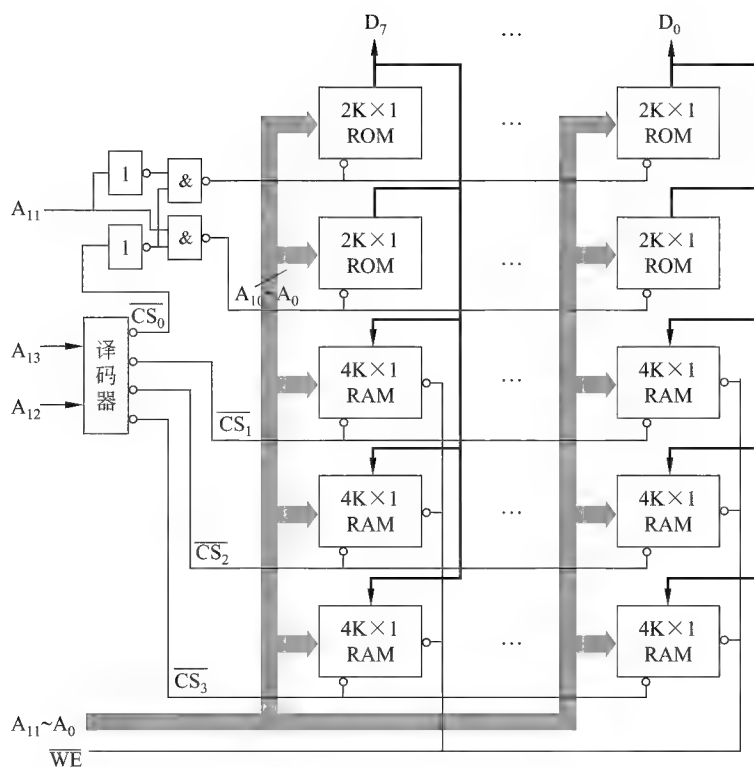


图 5 21 存储器的逻辑结构图

(2) 设存储器的读/写周期均为 $0.5\mu s$, CPU 在 $1\mu s$ 内至少要访存一次, 试问采用哪种刷新方式比较合理? 相邻两行之间的刷新间隔是多少? 对全部存储单元刷新一遍所需的实际刷新时间是多少?

解: (1) 存储器的结构框图如图 5-22 所示。

(2) 因为要求 CPU 在 $1\mu s$ 内至少要访存一次, 所以不能使用集中刷新方式, 分散和异步刷新方式都可以使用, 但异步刷新方式比较合理。

相邻两行之间的刷新间隔 = 最大刷新间隔时间 \div 行数 = $2ms \div 128 = 15.625\mu s$ 。取 $15.5\mu s$, 即进行读或写操作 31 次之后刷新一行。

对全部存储单元刷新一遍所需的实际刷新时间 = $0.5\mu s \times 128 = 64\mu s$

18. 有一个 8 位机, 采用单总线结构, 地址总线 16 位 ($A_{15} \sim A_0$), 数据总线 8 位 ($D_7 \sim D_0$), 控制总线中与主存有关的信号有 \overline{MREQ} (低电平有效允许访存) 和 R/\overline{W} (高电平为读命令, 低电平为写命令)。

主存地址分配如下: 从 $0 \sim 8191$ 为系统程序区, 由 ROM 芯片组成; 从 $8192 \sim 32767$

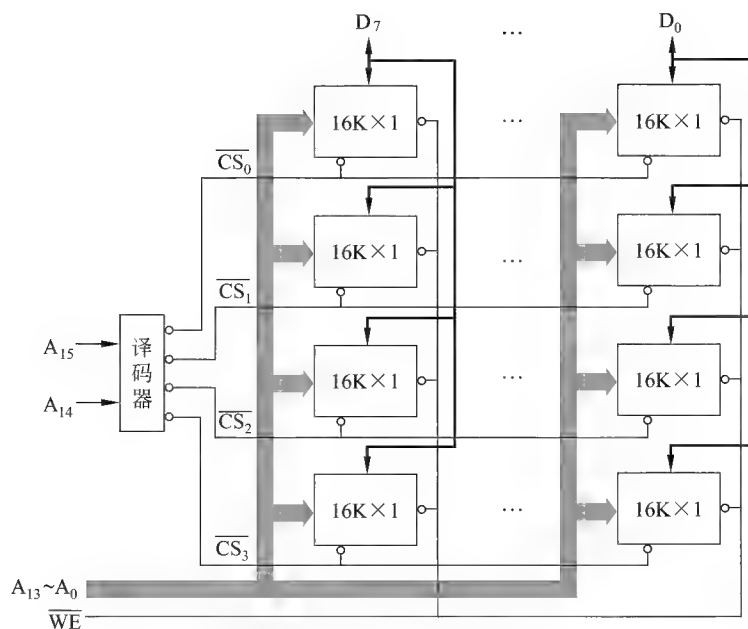


图 5 22 存储器的结构框图

为用户程序区;最后(最大地址)2K 地址空间为系统程序工作区。(上述地址均用十进制表示,按字节编址。)

现有如下存储芯片:8K×8 的 ROM,16K×1、2K×8、4K×8、8K×8 的 SRAM。请从上述规格中选用芯片设计该机主存储器,画出主存的连接框图,并注意画出片选逻辑及与 CPU 的连接。

解: 根据 CPU 的地址线、数据线,可确定整个主存空间为 64K×8。系统程序区由 ROM 芯片组成;用户程序区和系统程序工作区均由 RAM 芯片组成。共需:8K×8 的 ROM 芯片 1 片,8K×8 的 SRAM 芯片 3 片,2K×8 的 SRAM 芯片 1 片。

主存地址分配如图 5-23 所示,主存的连接框图如图 5-24 所示。

A ₁₅	A ₁₄	A ₁₃	A ₁₂	A ₁₁	A ₁₀ ~A ₀	
0	0	0	—	—	—	8KB ROM
0	0	1	—	—	—	8KB RAM
0	1	0	—	—	—	8KB RAM
0	1	1	—	—	—	8KB RAM
1	1	1	1	1	—	2KB RAM

19. 某半导体存储器容量 15KB,其中固化区 8KB,可选 EPROM 芯片为 4K×8;可随机

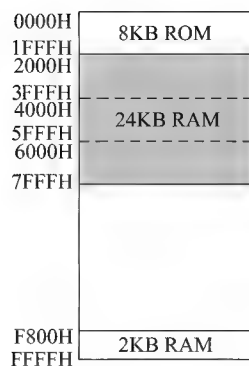


图 5-23 主存地址分配

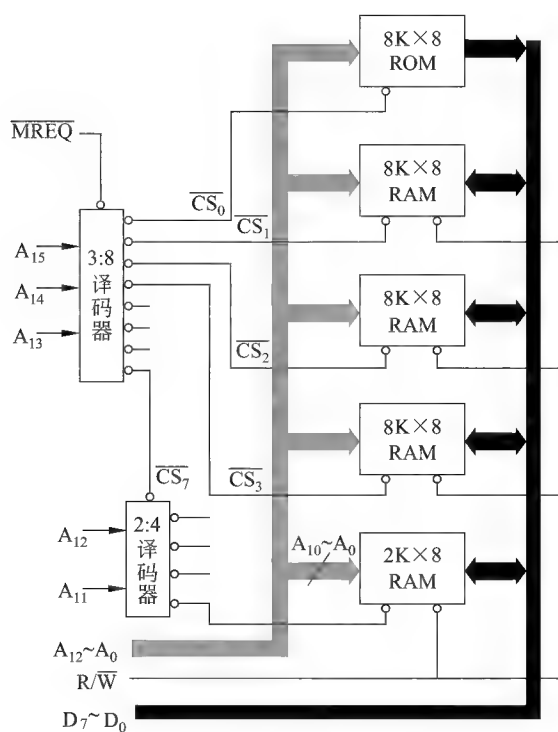


图 5-24 主存的连接框图

读/写区 7KB, 可选 SRAM 芯片有: $4K \times 4$ 、 $2K \times 4$ 、 $1K \times 4$ 。地址总线 $A_{15} \sim A_0$ (A_0 为最低位), 双向数据总线 $D_7 \sim D_0$ (D_0 为最低位), R/\overline{W} 控制读/写, \overline{MREQ} 为低电平时允许存储器工作信号。请设计并画出该存储器逻辑图, 注明地址分配、片选逻辑、片选信号极性等。

解：该存储器的地址分配如下：

4K×8 EPROM	0000H~0FFFH	8KB ROM
4K×8 EPROM	1000H~1FFFH	
4K×4 RAM (2片)	2000H~2FFFH	7KB RAM
2K×4 RAM (2片)	3000H~37FFH	
1K×4 RAM (2片)	3800H~3BFFH	

存储器逻辑图如图 5-25 所示。

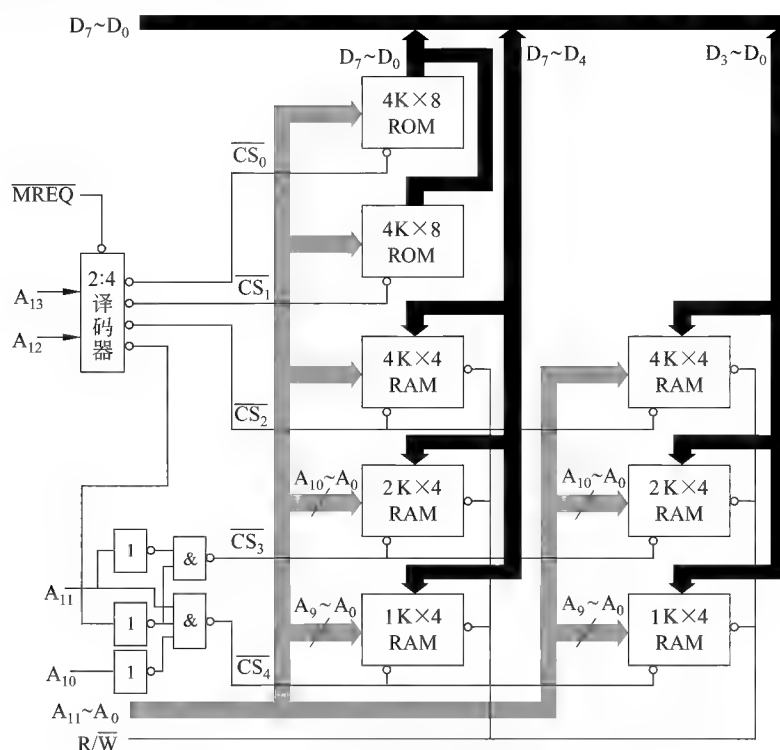


图 5 25 存储器逻辑图

假设采用部分译码方式,片选逻辑为:

$$\overline{CS_0} = A_{13} \cdot A_{12}$$

$$\overline{CS_1} = A_{13} \cdot A_{12}$$

$$\overline{CS_2} = A_{13} \cdot A_{12}$$

$$\overline{CS_3} = A_{13} \cdot A_{12} \cdot A_{11}$$

$$\overline{CS_4} = A_{13} \cdot A_{12} \cdot A_{11} \cdot A_{10}$$

20. 某机地址总线 16 位 $A_{15} \sim A_0$ (A_0 为最低位), 访存空间 64KB。外围设备与主存统一编址, I/O 空间占用 $FC00 \sim FFFFH$ 。现用 2164 芯片 ($64K \times 1$) 构成主存储器, 请设计并画出该存储器逻辑图, 并画出芯片地址线、数据线与总线的连接逻辑以及行选信号与列选信号的逻辑式, 使访问 I/O 时不访问主存。动态刷新逻辑可以暂不考虑。

解: 存储器逻辑图如图 5-26 所示, 为简单起见, 在图中没有考虑行选信号和列选信号, 行选信号和列选信号的逻辑式可参考下题。

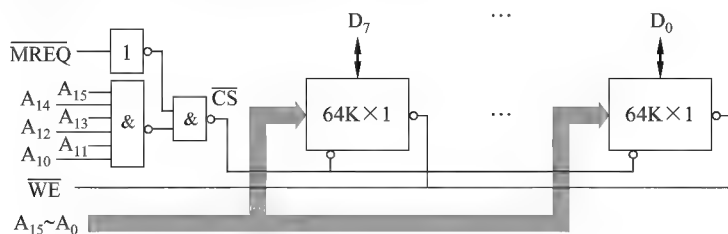


图 5-26 存储器逻辑图

在 64KB 空间的最后 1KB 为 I/O 空间, 在此区间 \overline{CS} 无效, 不访问主存。

21. 已知有 $16K \times 1$ 的 DRAM 芯片, 其引脚功能如下: 地址输入 $A_6 \sim A_0$, 行地址选择 \overline{RAS} , 列地址选择 \overline{CAS} , 数据输入端 D_{IN} , 数据输出端 D_{OUT} , 控制端 \overline{WE} 。请用给定芯片构成 256KB 的存储器, 采用奇偶校验, 试问: 需要芯片的总数是多少? 并请:

- (1) 正确画出存储器的连接框图。
- (2) 写出各芯片 \overline{RAS} 和 \overline{CAS} 形成条件。
- (3) 若芯片内部采用 128×128 矩阵排列, 求异步刷新时该存储器的刷新闻隔。

解: (1) 需要的芯片数 = 128 片, 存储器的连接框图如图 5-27 所示。

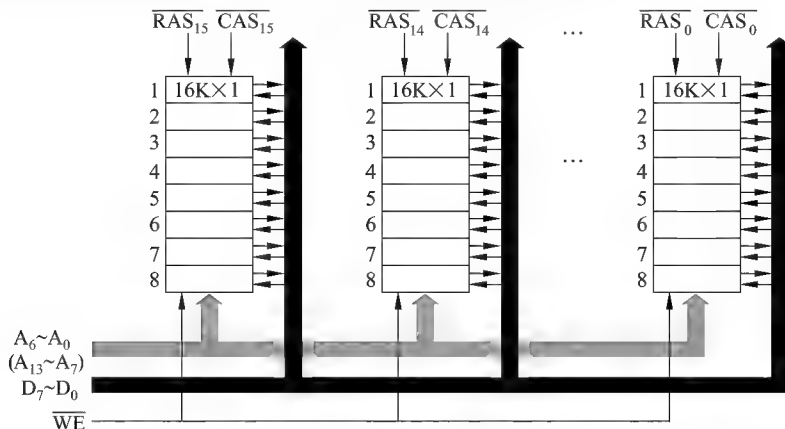


图 5-27 存储器的连接框图

(2) 存储器正常读写操作时, RAS比CAS先有效, 由于行、列分时传送, 所以RAS与CAS也应分时出现, 且RAS在先, CAS在后, 分别与时间因素 t_1 , t_2 有关。 $A_{17} \sim A_{14}$ 用于译码选择 16 个不同的 16KB 空间, 译码电路如图 5-28 所示, $\overline{\text{RAS}}$ 和 $\overline{\text{CAS}}$ 的形成条件分别为:

$$\overline{\text{RAS}}_0 = A_{17} \cdot A_{16} \cdot A_{15} \cdot A_{14} \cdot t_1$$

...

$$\overline{\text{RAS}}_{15} = A_{17} \cdot A_{16} \cdot A_{15} \cdot A_{14} \cdot t_1$$

$$\overline{\text{CAS}}_0 = A_{17} \cdot A_{16} \cdot A_{15} \cdot A_{14} \cdot t_2$$

...

$$\overline{\text{CAS}}_{15} = A_{17} \cdot A_{16} \cdot A_{15} \cdot A_{14} \cdot t_2$$

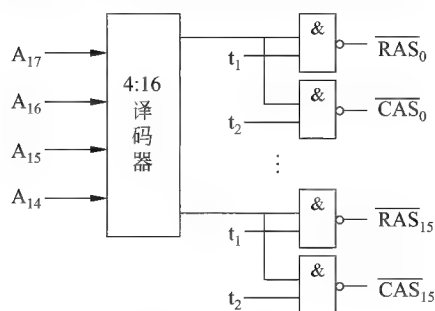


图 5 28 译码电路

(3) 若芯片内部采用 128×128 矩阵排列, 设芯片的最大刷新闻隔时间为 2ms , 则相邻两行之间的刷新闻隔为:

$$\text{刷新闻隔} = \text{最大刷新闻隔时间} \div \text{行数} = 2\text{ms} \div 128 = 15.625\mu\text{s}$$

可取刷新闻隔 $15.5\mu\text{s}$ 。

22. 并行存储器有哪几种编址方式? 简述低位交叉编址存储器的工作原理。

解: 并行存储器有单体多字、多体单字和多体多字等几种系统。

多体交叉访问存储器可分为高位交叉编址存储器和低位交叉编址存储器。低位交叉编址又称为横向编址, 连续的地址分布在相邻的存储体中, 而同一存储体内的地址都是不连续的。存储器地址寄存器的低位部分经过译码选择不同的存储体, 而高位部分则指向存储体内的存储字。如果采用分时启动的方法, 可以在不改变每个存储体存取周期的前提下, 提高整个主存的速度。

23. 什么是高速缓冲存储器? 它与主存是什么关系? 其基本工作过程如何?

解: 高速缓冲存储器位于主存和 CPU 之间, 用来存放当前正在执行的程序段和数据

中的活跃部分,使 CPU 的访存操作大多数针对 Cache 进行,从而使程序的执行速度大大提高。

高速缓冲存储器的存取速度接近于 CPU 的速度,但是容量较小,它保存的信息只是主存中最急需处理的若干块的副本。

当 CPU 发出读请求时,如果 Cache 命中,就直接对 Cache 进行读操作,与主存无关;如果 Cache 不命中,则仍需访问主存,并把该块信息一次从主存调入 Cache 内。若此时 Cache 已满,则须根据某种替换算法,用这个块替换掉 Cache 中原来的某块信息。

24. Cache 做在 CPU 芯片内有什么好处? 将指令 Cache 和数据 Cache 分开又有什么好处?

解: Cache 做在 CPU 芯片内可以提高 CPU 访问 Cache 的速度。将指令 Cache 和数据 Cache 分开的好处是分体缓存支持并行访问,即在取指部件取指令的同时,取数部件要取数据。并且,指令在程序执行中一般不需要修改,故指令 Cache 中的内容不需写回到主存中去。

25. 设某机主存容量为 4MB,Cache 容量为 16KB,每块包含 8 个字,每字 32 位,设计一个四路组相联映像(即 Cache 每组内共有四个块)的 Cache 组织,要求:

(1) 画出主存地址字段中各段的位数。

(2) 设 Cache 的初态为空,CPU 依次从主存第 0、1、2、…、99 号单元读出 100 个字(主存一次读出一个字),并重复按此次序读 8 次,问命中率是多少?

(3) 若 Cache 的速度是主存的 6 倍,试问有 Cache 和无 Cache 相比,速度提高多少倍?

解: (1) 主存容量为 4MB,按字节编址,所以主存地址为 22 位,地址格式如图 5-29 所示。

区号 (8 位)	组号 (7 位)	组内块号 (2 位)	块内地址 (5 位)
-------------	-------------	---------------	---------------

图 5 29 主存地址格式

(2) 由于每个字块有 8 个字,所以主存第 0、1、2、…、99 号字单元分别在字块 0~12 中,采用四路组相联映像将分别映像到第 0 组~12 组中,但 Cache 起始为空,所以第一次读时每一块中的第一个单元没命中,但后面 7 次每个单元均可以命中。

$$\text{命中率} = \frac{N_c}{N_c + N_m} = \frac{100 - 13 + 7 \times 100}{8 \times 100} = 98.4\%$$

(3) 设 Cache 的存取周期为 T ,则主存的存取周期为 $6T$ 。

有 Cache 的访存时间 = $H \times T_c + (1 - H) \times (T_m + T_c) = T_c + (1 - H) \times T_m = T + (1 - 98.4\%) \times 6T = 1.096T$

无 Cache 的访存时间为 $6T$

所以速度提高倍数 $= 6 \div 1.096 = 5.47$ 倍。

26. 什么叫虚拟存储器? 采用虚拟存储技术能解决什么问题?

解: 虚拟存储器由主存储器和联机工作的辅助存储器(通常为磁盘存储器)共同组成,这两个存储器在硬件和系统软件的共同管理下工作,对于应用程序员,可以把它们看作是一个单一的存储器。

采用虚拟存储技术可以解决主存容量不足的问题。虚拟存储器将主存和辅存的地址空间统一编址,形成一个庞大的存储空间。在这个大空间里,用户可以自由编程,完全不必考虑程序在主存是否装得下以及这些程序将来在主存中的实际存放位置。

27. 已知采用页式虚拟存储器,某程序中一条指令的虚地址是: 000001111111100000。该程序的页表起始地址是 0011,页面大小 1K,页表中有关单元最末四位(实页号)见下表:

虚页号	装入位	实页号
007H	1	0001
...
300H	1	0011
...
307H	1	1100

请指出指令地址(虚地址)变换后的主存实地址。

解: 页面大小 1K,页内地址 10 位,根据页表,可以得出主存实地址为 11001111100000。

第 6 章

中央处理器

6.1 基本内容要求

中央处理器(CPU)是整个计算机的核心,它包括运算器和控制器。本章着重讨论 CPU 的功能和组成、控制器的工作原理和实现方法、微程序控制原理、基本控制单元的设计以及先进的流水线技术和 RISC 技术。

学习要求

- ◆ 理解 CPU 的功能
- ◆ 理解 CPU 中的通用寄存器和专用寄存器的设置及作用
- ◆ 了解 CPU 的主要技术参数
- ◆ 了解控制器的基本组成
- ◆ 理解微操作信号发生器输出信号的产生
- ◆ 掌握组合逻辑控制器和微程序控制器的区别
- ◆ 理解时序系统中指令周期、机器周期的概念
- ◆ 理解不同的控制方式(同步、异步、联合)
- ◆ 理解一条指令执行的基本过程
- ◆ 掌握取指周期的微操作序列(公共操作)
- ◆ 理解微程序控制的有关术语(如微命令、微操作、微指令、微程序等)
- ◆ 理解微程序控制计算机的两个层次(传统机器层和微程序层)
- ◆ 掌握各种微指令编码法的特点,微指令操作控制字段的设计
- ◆ 理解微程序控制器的组成,熟悉其特有部件的作用
- ◆ 了解微程序控制器执行指令的工作过程
- ◆ 掌握微程序入口地址的形成方法
- ◆ 理解微程序后继微地址的形成方法
- ◆ 了解微程序设计技术

- ◆ 了解组合逻辑控制单元的设计
- ◆ 了解微程序控制单元的设计
- ◆ 了解流水线技术
- ◆ 了解 RISC 的特点和基本技术

6.2 误点疑点解惑

1. CPU 中寄存器的设置

CPU 中有许多寄存器,一般将它们分为通用寄存器和专用寄存器两大类,也有些书上将它们细分为用于处理的寄存器、用于控制的寄存器和用于主存接口的寄存器。

(1) 通用寄存器

通用寄存器也就是用于处理的寄存器,它们可提供操作数并存放运算结果,或作为地址指针,或作为基址寄存器、变址寄存器,或作为计数器等。在指令系统中为这些寄存器分配了编号,可以编程指定使用其中的某个寄存器,对程序员来说是“看得见”的寄存器。

在对这组寄存器的设计上,有的计算机将它们设计成基本通用,如 PDP-11 中的通用寄存器命名为 R_0 、 R_1 、 R_2 、...,它们可被指定担任各种工作,大部分寄存器没有特定的任务上的分工。有的计算机则为这组寄存器分别规定了某一基本任务,并按各自的基本任务命名,如 Intel 80x86 中设置有累加器 AX、基址寄存器 BX、数据寄存器 DX 等。

CPU 中还常设置一些用户不能直接访问的寄存器组用来暂存信息,称为暂存器。在指令系统中没有为它们分配编号,因而不能直接编程访问。对程序员来说,它们是看不见的。

(2) 专用寄存器

CPU 至少有 5 个专用寄存器,它们又被分为用于控制的寄存器和用于主存接口的寄存器。

用于控制的寄存器在 CPU 中起着控制操作的作用,控制寄存器有:

程序计数器(PC) 又称为指令指针(IP),用来存放指令地址。为了保证程序能自动连续执行,CPU 必须能自动确定下一条指令的地址。在程序开始执行前,将程序的第一条指令所在的存储单元地址送入 PC,以便从程序的第一条指令开始执行。在程序执行过程中,CPU 将自动修改 PC 的内容,使其保存的总是将要执行的下一条指令的地址。顺序执行时,PC 增量计数(加“1”);遇到转移指令,则将转移地址送至 PC。需要提醒学生注意的是,所谓 PC 加“1”中的“1”,代表的是一条指令,而不一定是一个存储单元,这取决于主存的编址方式,若主存按字节编址,则增量值为指令所占的字节数,即指令占 1 个字节,PC+1;指令占 2 个字节,PC+2;……。

指令寄存器(IR) 用来存放现行指令。当执行一条指令时,首先从主存将指令取出

送到指令寄存器中去,直到这一条指令执行完毕再放入下条指令。为了提高指令的执行速度,现在大多数计算机都将指令寄存器扩充为指令队列(指令栈),允许预取若干条指令。

状态标志寄存器(PSWR) 用来存放程序状态字,其内容表示现执行程序的状态。一条指令执行完毕后,除去结果存放于寄存器或存储器之外,还将根据运行结果自动修改标志位的有关内容,这些内容可被后面的条件转移指令所测试,作为决定程序流向的因素之一。

主存接口寄存器用于 CPU 与主存储器的数据交换,主存接口寄存器有:

存储器地址寄存器(MAR) 用来接收指令地址(PC 的内容)、操作数地址或结果数据地址,以确定要访问的单元。

存储器数据寄存器(MDR) 也可称为存储器数据缓冲寄存器(MBR)。写入主存的数据一般先送至 MDR,再送入主存;从主存读出的指令或数据一般先送入 MDR,再送入指定寄存器。

2. CPU 的主要性能指标参数

(1) 主频

主频是 CPU 内核(整数和浮点运算器)电路的实际运行频率。一般来说,主频越高,CPU 在一个时钟周期里所能执行的指令数也就越多,CPU 的运算速度也就越快。

通常,同一类型计算机在 1 个时钟周期内执行的平均指令数是固定的,如 Pentium 在 1 个时钟周期内执行 2 条运算指令。那么主频为 100MHz 的 Pentium 可以在 1 秒钟内执行 2 亿条指令,主频为 200MHz 的 Pentium 可以在 1 秒钟内执行 4 亿条指令,主频为 1GHz 的 Pentium 则可以在 1 秒钟内执行 20 亿条指令。

主频无疑是显示 CPU 性能的最根本的指标,但若认为 CPU 的主频指的就是 CPU 运行速度则太片面。CPU 的主频表示在 CPU 内数字脉冲信号震荡的速度,从根本上讲,与 CPU 实际的运算能力是没有直接关系的,在一定情况下,很可能会出现主频较高的 CPU 实际运算速度较低的现象。

(2) 外频

外频是 CPU 总线频率,是主板为 CPU 提供的基准频率。外频是指 CPU 与主板之间同步运行的速度,而且目前的绝大部分电脑系统中外频也是主存与主板之间同步运行的速度,在这种方式下,可以理解为 CPU 的外频直接与主存相连通,实现两者间的同步运行状态。

外频提高后,CPU 与主板之间的交换速度也相应得到了提高,这对提高计算机整体运行速度影响较大。

(3) 倍频

倍频就是 CPU 的运行频率与整个系统外频运行频率之间的倍数,即主频=外频×

倍频。在相同的外频下,倍频越高,CPU的主频也越高。但实际上,在相同外频的前提下,高倍频的CPU本身意义并不大。这是因为前面提到CPU与系统之间数据传输速度是有限的,单纯的一味追求高倍频而得到高主频的CPU就会出现明显的“瓶颈”效应——CPU从系统中得到数据的极限速度不能满足CPU运算的速度,可想而知,这样无疑是一种浪费。通常,倍频值为几倍到十几倍。

(4) 前端总线频率

前端总线(front side bus,FSB)频率是目前微机系统中经常涉及的一个性能指标参数,在某些微机中,它与外频是一回事,但在另一些微机中,它们又有所不同,详见本章其后相关知识介绍中的内容。

前端总线频率将影响计算机运行时CPU与主存之间的数据交换速度,实际上也就影响了计算机的整体运行速度。

(5) L1 Cache

一级高速缓存,其容量一般为几十~几百KB,频率与CPU相同。L1 Cache可以提高CPU的运行效率,它的容量和结构对CPU的性能影响较大。L1 Cache越大,系统性能提高越明显。

(6) L2 Cache

二级高速缓存,其容量一般相当于L1 Cache的4~8倍。L2 Cache的容量和频率对CPU的性能影响也较大,L2 Cache的时钟频率一般为CPU主频的 $1/2$ 、 $2/5$ 和 $1/3$ 。

(7) 地址总线宽度

地址总线宽度决定了CPU可以访问的存储器物理地址空间,这里的物理空间的大小是指主存容量,因为从硬盘等外部存储器中来的数据必须经过主存才能得到CPU的访问。例如,地址总线宽度为32位,表示CPU最多可以直接访问4GB的物理空间,而硬盘等外部存储器的容量则不受此限制。

(8) 数据总线宽度

数据总线宽度则决定了CPU与外部Cache、主存以及输入输出设备之间进行一次数据传输的信息量。

3. 控制器的功能与组成

控制器是整个计算机的指挥中心,为保证机器有条不紊的工作,主要完成如下功能。

① 指令控制功能:计算机的工作过程是连续执行指令的过程,指令在主存中连续存放,一般情况下,指令被顺序执行,只有遇到转移指令才会改变指令的执行顺序,所以指令在主存中的存放顺序是静态的,而指令的执行顺序(指令流)是动态的,控制器应能保证指令流的正常流动。

② 时序控制功能:由于各条机器指令的复杂长度不同,所以每个指令周期中包含的机器周期数各不相同,各个机器周期中包含多少个节拍也不一定相同,控制器必须产生指

令周期、机器周期和节拍等时序信号,用来给机器定时。

③ 操作控制功能:在时序信号的控制下,每条机器指令在各个机器周期的各个节拍中应产生哪些微操作控制信号是有严格规定的,控制器应根据指令的操作流程,在各个节拍中产生相应的微操作控制信号,以有效地完成各条指令的操作过程。

控制器主要由以下几部分组成:

- ① 指令部件。包括程序计数器、指令寄存器、指令译码器、地址形成部件等。
- ② 时序部件。包括脉冲源、启停控制逻辑、节拍信号发生器等。
- ③ 微操作信号发生器。
- ④ 中断控制逻辑。

需要特别注意的是,暂存在指令寄存器中的指令,其操作码部分经译码后才能识别出当前要执行的指令是一条什么样的指令,也就是说,指令译码器仅对指令中的操作码字段进行译码,而不是对整条指令进行译码。

4. 控制器的核心——控制单元

控制器的核心是微操作信号发生器(控制单元 CU),计算机无论执行什么任务,都是在控制单元的控制下完成的。控制单元通常有 3 种实现方法:组合逻辑电路、存储逻辑电路、可编程逻辑阵列 PLA。根据控制单元实现方法的不同,控制器可分为组合逻辑型、存储逻辑型、组合逻辑与存储逻辑结合型 3 种。这 3 种不同类型的控制器仅仅是控制单元实现不同,而控制器中的其他部分基本上是大同小异的。

控制单元的输入包括时序信号、机器指令操作码、各部件状态反馈信号等,输出的微操作控制信号又可以细分为 CPU 内的控制信号和送至主存或外设的控制信号。CPU 内部的控制信号用于控制寄存器间的数据传送,使 ALU 完成指定的功能以及其他的内部操作。发送至 CPU 外部的控制信号用于控制 CPU 与主存和外设交换数据。

控制单元的一般模型如图 6-1 所示,该模型表示了控制单元的输入和输出信号之间的关系。

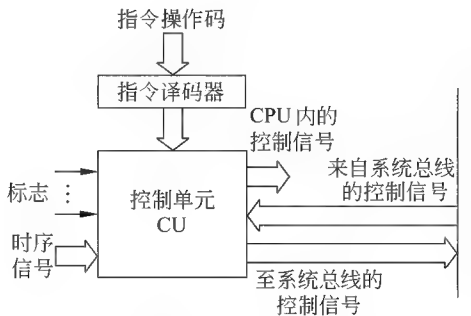


图 6-1 控制单元模型

控制单元的输入信号主要有：

- ① 时序信号。CPU 的所有工作都必须按一定的时间关系有序的安排。
- ② 指令译码器输出结果。当前指令的操作码译码之后用于确定该指令应该完成何种微操作。
- ③ 标志。控制单元需要一些标志来决定 CPU 应该发出哪些控制信号,例如,对“增量若为 0 跳步(ISZ)”指令来说,控制单元将根据零标志是否置位来确定 PC 是否加“1”。
- ④ 来自系统总线的控制信号。中断信号和存储器操作完成信号等。

控制单元的输出信号主要有：

- ① CPU 内的控制信号。包括用于寄存器之间传送数据和用于指定 ALU 功能的两类控制信号。
- ② 至控制总线的控制信号。包括对存储器的控制信号和对外设的控制信号。

5. 指令的机器周期

一条指令从读取到执行完的全部时间称为指令周期。通常,每个指令周期中采用机器周期、节拍、工作脉冲三级时序系统。

不同类型指令所需的机器周期数可能不同,一条指令至少需要两个机器周期。通常,有 4 个机器周期用于指令的正常执行,另外还有两个机器周期(中断、DMA)用于 I/O 传送控制。根据各类指令执行的需要,为每个机器周期设置一个触发器作为标志。某一时期内有一个且仅有一个触发器置“1”,以此来指明 CPU 现在所处的执行指令的阶段。下面讨论几个常见的机器周期。

① 取指周期 FT:取指周期完成取指令的工作,这是每条指令都必须经历的。在 FT 中完成的操作与指令的操作码无关,但取指周期结束后将转向哪个机器周期,则与 FT 中取出的指令类型及所采用的寻址方式有关。

② 取源操作数周期 ST:如果需要从主存中读取源操作数(非寄存器寻址),则进入 ST。在 ST 中将根据指令的源地址字段形成源操作数地址,读取源操作数。

③ 取目的操作数周期 DT:如果需要从主存中读取目的操作数(非寄存器寻址),则进入 DT。在 DT 中将根据指令的目的地址字段形成目的操作数地址,读取目的操作数。

④ 执行周期 ET:这是各类指令都需经历的最后一个工作阶段,在 ET 中将根据指令的操作码执行相应的操作,如传送、算术运算、逻辑运算、保存返回地址、获得转移地址等。

⑤ 中断周期 IT:除了考虑指令的正常执行,还需考虑外部请求带来的变化。在响应中断请求之后,到执行中断服务程序之前,需要一个过渡期,这就是中断周期 IT。在 IT 中直接依靠硬件进行关中断、保护断点、转中断服务程序入口等操作。

⑥ DMA 周期 DMAT:需要 DMA 请求之后,CPU 进入 DMAT。在 DMAT 中,CPU 交出系统总线的控制权,改由 DMA 控制器控制系统总线,实现主存与外设间的数据直接传送。因此对 CPU 来说,DMAT 是一个空操作周期。各机器周期状态之间的转换如图

6-2 所示。

FT 结束后,对于双操作数指令,如果数均在主存中,则先进入 ST,之后进入 DT、ET;如果数均在寄存器中,则进入 ET;对于单操作数指令,如果数在主存中,则进入 DT、ET,如果数在寄存器中,同样进入 ET;对于转移指令,FT 结束后直接进入 ET。因此,在每一机器周期结束前,都要判断下一个周期状态将是什么,并且准备进入下一周期的条件。到本周期结束的时刻,再实现周期状态的定时切换。

由于 DMA 周期实现的是高速数据传送,所以让 DMA 请求的优先级高于中断请求。因而在一条指令将要结束时,先判断有无 DMA 请求,若有请求,将插入 DMAT。请注意,实际上计算机允许在每个机器周期结束时就插入 DMAT,但为简化控制逻辑,在图 6-2 中限制在一条指令结束时才判别与响应 DMA 请求。如果在一个 DMAT 结束前又提出新的 DMA 请求,则允许连续安排若干个 DMA 周期。

若没有 DMA 请求,则判断有无中断请求。若有中断请求,则进入 IT,在 IT 中完成必要的过渡期工作后,将转向新的 FT,开始读取中断服务程序的第一条指令;如果没有中断请求,就返回 FT,开始读取现行程序的后续指令。

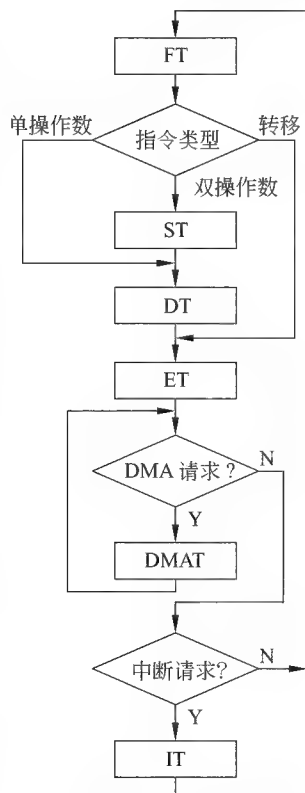


图 6 2 各机器周期之间的转换

以上的许多机器周期中既有 CPU 内部数据通路操作,也有访问主存的操作,为了简化时序控制,令机器周期就等于主存的存取周期。这对于 CPU 内部操作来说,无疑在时间上是比较浪费的。

6. 指令执行的控制方式

在控制器设计时,可以采用同步控制、异步控制和联合控制这 3 种方式来实现指令执行时间的控制或不同部件之间的控制。下面以指令执行控制为例讨论各种控制方式。

(1) 同步控制方式

同步控制方式的特点可以归纳为以下 3 点:

- ① 以微操作序列最长的指令为标准,确定控制微操作运行的节拍数;
- ② 控制器产生统一的、顺序固定的、周而复始的节拍电位和工作脉冲;
- ③ 简单指令(微操作序列短的指令)可空着一部分节拍不用。也就是说,不管什么指令,实现的时间都是相同的。

同步控制方式又称作中央控制方式,其优点是控制电路简单,缺点是运行速度较慢。

(2) 异步控制方式

异步控制方式的特点可以归纳为 3 点：

- ① 每条指令需要多少节拍,就产生多少节拍；
- ② 指令执行完毕,发出回答信号；
- ③ 控制器收到回答信号即开始执行下条指令。

异步控制方式又称作局部控制方式,其优点是运行速度快,缺点是控制电路比较复杂。

(3) 联合控制方式

联合控制方式是把同步控制方式和异步控制方式结合使用的一种控制方式,又可称为混合控制方式,它的特点可归纳为两点：

- ① 大部分指令按同步方式执行；
- ② 小部分特殊指令(微操作序列过长或微操作时间难以确定的指令)采用异步控制方式执行。

这种方式下,大多数指令都采用相同的节拍(中央控制);对于某些指令,如乘法、除法、移位等所需运算时间较长的指令,则采用异步控制(局部控制)。局部控制周期的长度可以根据指令对应操作步的具体需要而定。在局部控制周期结束时,再次进入中央控制,完成指令处理的所有操作步骤。联合控制的处理过程如图 6-3 所示。

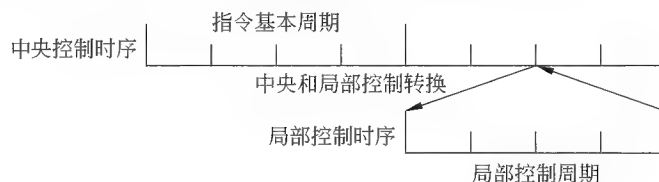


图 6-3 联合控制的处理过程

7. 指令微操作序列的安排

控制器在实现一条指令的功能时,总要把每条指令分解成为一系列时间上先后有序的最基本、最简单的微操作,即微操作序列。指令操作流程与相应微操作序列的安排,主要取决于数据通路的结构,不同的数据通路有不同的微操作序列。某一假想机的数据通路如图 6-4 所示,该机采用单总线结构,CPU、主存和外设都挂在总线上。图 6-4 中箭头表示信息传送方向。“□”表示控制门,上面标有控制信号。当某一控制信号为高电平时,控制门被打开,允许一次信息流动。当控制信号为低电平时,控制门被关闭,信息不能流动。

假想机的运算部件以 ALU 为核心,两个输入端只设置了一个锁存器 LA,另一个输入端是来自总线的数据,输出直接送暂存器 LT。暂存器 TEMP 只用于在指令执行过程存储数据,对用户是完全透明的。程序计数器 PC 通过 ALU 实现+1 操作。图 6-4 中标

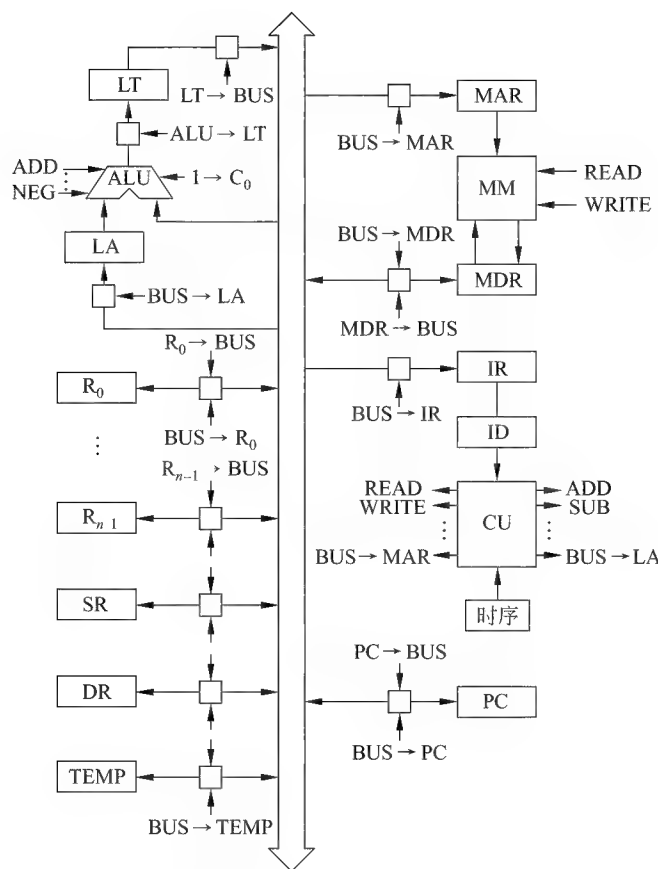


图 6 4 假想机的数据通路

出的控制信号,即为微操作控制信号(控制计算机的最简单不能再分解的控制信号),它实际控制数据通路中的数据流和指令流的流向。这些控制信号在本质上是控制数据通路的各个控制门的打开或关闭、ALU 的实际操作、寄存器接数、主存的读或写等。这些微操作是有时序的,何时发生,有哪些,完全由指令的功能(IR 中的操作码字段)决定。除此之外,还有 CLEAR LA 以及 ALU 的功能控制信号等。

假设机器设置 4 个机器周期,一个机器周期包含 4 个节拍,节拍数固定,对于操作步骤较少的指令,有些节拍可能轮空,时间上有些浪费。

每条指令的取指周期(FT)完成的任务是相同的,相应的微操作序列如下:

T_1 PC→BUS, BUS→MAR, READ, CLEAR LA, $1 \rightarrow C_0$, ADD, ALU→LT

T_2 LT→BUS, BUS→PC

T_3 MDR→BUS, BUS→IR

T_4 $1 \rightarrow ST$

下面介绍几条有代表性的指令的微操作序列,图 6-5 为这几条指令的操作流程图。

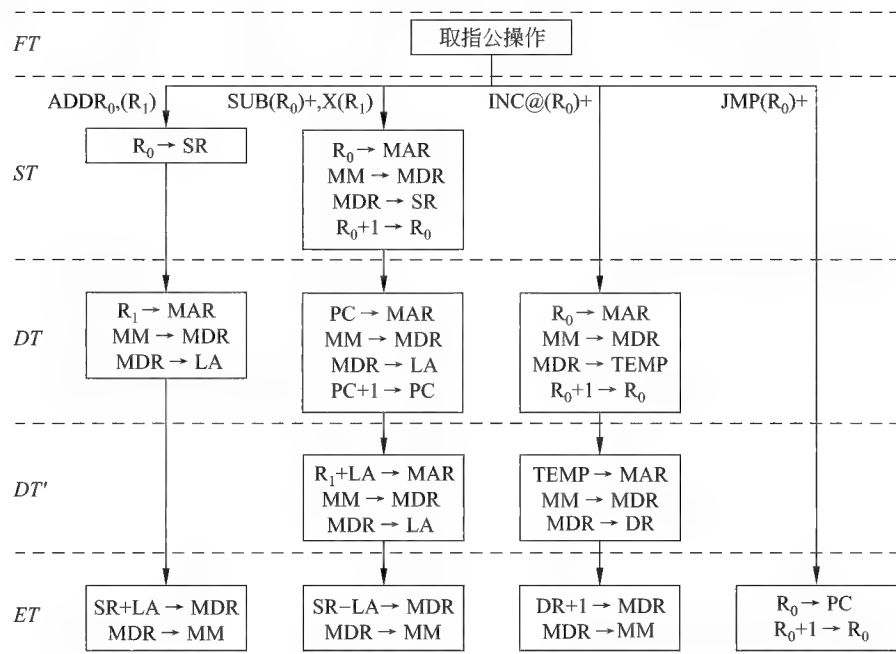


图 6-5 指令的操作流程图

例 6-1 写出加法指令 $ADD R_0, (R_1)$ 的微操作序列。

解:

START

FT 微操作序列同上

ST

T_1 $R_0 \rightarrow BUS, BUS \rightarrow SR$

T_2 空操作

T_3 空操作

T_4 $1 \rightarrow DT$

DT

T_1 $R_1 \rightarrow BUS, BUS \rightarrow MAR, READ$

T_2 $MDR \rightarrow BUS, BUS \rightarrow LA$

T_3 空操作

T_4 $1 \rightarrow ET$

ET

T_1 $SR \rightarrow BUS, ADD, ALU \rightarrow LT$

T_2 $LT \rightarrow BUS, BUS \rightarrow MDR, WRITE$

T_3 空操作

T_4 END

例 6-2 写出减法指令 $SUB(R_0)+, X(R_1)$ 的微操作序列。

解:

START

FT 微操作序列同上

ST

T_1 $R_0 \rightarrow BUS, BUS \rightarrow MAR, READ, CLEAR LA, 1 \rightarrow C_0, ADD, ALU \rightarrow LT$

T_2 $LT \rightarrow BUS, BUS \rightarrow R_0,$

T_3 $MDR \rightarrow BUS, BUS \rightarrow SR$

T_4 $1 \rightarrow DT$

DT

T_1 $PC \rightarrow BUS, BUS \rightarrow MAR, READ, CLEAR LA, 1 \rightarrow C_0, ADD, ALU \rightarrow LT$

T_2 $LT \rightarrow BUS, BUS \rightarrow PC$

T_3 $MDR \rightarrow BUS, BUS \rightarrow LA$

T_4 $1 \rightarrow DT'$

DT'

T_1 $R_1 \rightarrow BUS, ADD, ALU \rightarrow LT$

T_2 $LT \rightarrow BUS, BUS \rightarrow MAR, READ$

T_3 $MDR \rightarrow BUS, BUS \rightarrow LA$

T_4 $1 \rightarrow ET$

ET

T_1 $SR \rightarrow BUS, SUB, ALU \rightarrow LT$

T_2 $LT \rightarrow BUS, BUS \rightarrow MDR, WRITE$

T_3 空操作

T_4 END

此例中,源操作数采用自增型间接寻址,以 R_0 为地址访问主存一次,从主存中取出源操作数送入源操作数寄存器 SR ,并使 R_0 的内容+1。目的操作数采用变址寻址,指令的第二个字是形式地址(位移量)。在 DT 周期中,先以 PC 现行值为地址从存储单元取得位移量 X ,再与 R_1 的内容相加,以相加结果为地址取出操作数送入锁存器 LA ,同时 PC 的内容+1,准备好下条指令地址。由于取目的操作数阶段需要两次访问主存,所以 DT

周期必须重复一次。

例 6-3 写出加 1 指令 $\text{INC} @ (R_0) +$ 的微操作序列。

解：

START

FT 微操作序列同上 ($T_3 \quad 1 \rightarrow DT$)

DT

$T_1 \quad R_0 \rightarrow \text{BUS}, \text{BUS} \rightarrow \text{MAR}, \text{READ}, \text{CLEAR LA}, 1 \rightarrow C_0, \text{ADD}, \text{ALU} \rightarrow \text{LT}$

$T_2 \quad \text{LT} \rightarrow \text{BUS}, \text{BUS} \rightarrow R_0,$

$T_3 \quad \text{MDR} \rightarrow \text{BUS}, \text{BUS} \rightarrow \text{TEMP}$

$T_4 \quad 1 \rightarrow DT'$

DT'

$T_1 \quad \text{TEMP} \rightarrow \text{BUS}, \text{BUS} \rightarrow \text{MAR}, \text{READ}$

$T_2 \quad \text{MDR} \rightarrow \text{BUS}, \text{BUS} \rightarrow \text{DR}$

$T_3 \quad$ 空操作

$T_4 \quad 1 \rightarrow \text{ET}$

ET

$T_1 \quad \text{DR} \rightarrow \text{BUS}, \text{CLEAR LA}, 1 \rightarrow C_0, \text{ADD}, \text{ALU} \rightarrow \text{LT}$

$T_2 \quad \text{LT} \rightarrow \text{BUS}, \text{BUS} \rightarrow \text{MDR}, \text{WRITE}$

$T_3 \quad$ 空操作

$T_4 \quad \text{END}$

此例中,目的操作数采用自增型双间址, R_0 的内容是操作数地址的地址,第一次访问主存取出的是操作数的地址,送入存储器地址寄存器 MAR ,以此地址再访问主存取出的是操作数,送入目的操作数寄存器 DR 。除此之外, R_0 的内容+1。在取目的操作数阶段需要两次访问主存,所以 *DT* 周期要重复一次。

例 6-4 写出转移指令 $\text{JMP} (R_0) +$ 的微操作序列。

解：

START

FT 微操作序列同上 ($T_3 \quad 1 \rightarrow \text{ET}$)

ET

$T_1 \quad R_0 \rightarrow \text{BUS}, \text{BUS} \rightarrow \text{PC}, \text{CLEAR LA}, 1 \rightarrow C_0, \text{ADD}, \text{ALU} \rightarrow \text{LT}$

$T_2 \quad \text{LT} \rightarrow \text{BUS}, \text{BUS} \rightarrow R_0,$

$T_3 \quad$ 空操作

$T_4 \quad \text{END}$

以上是 4 条有代表性的指令的微操作序列,上述安排并不是最优的方案。事实上,指

令的微操作序列是机器所有指令的微操作在各个时序信号上的分配,它是指令流程的进一步具体化。安排微操作序列必须遵循两个简单的原则:

① 微操作序列的顺序必须是恰当的。例如,必须保证 $PC \rightarrow BUS, BUS \rightarrow MAR$ 信号先于 $MDR \rightarrow BUS, BUS \rightarrow IR$ 信号,因为存储器读操作需使用 MAR 地址。

② 不能引起数据通路上的信息发生冲突。例如,在一个节拍内不能两次往总线发送信息。

安排好每一条指令的微操作序列是一个比较复杂的问题,初学者感觉困难较大。建议在教学过程中不要太多的关注细节问题,主要从指令的几个机器周期出发,讲清楚每个机器周期应当做些什么,如何做就可以了。

8. 字段直接编码和字段间接编码

在控制数据通路的操作中,大多数微命令是不会同时出现的,例如,控制 ALU 操作的各种微命令(如 ADD、SUB、AND 等)就不能同时出现,即在一条微指令中只能同时出现一种运算操作;又例如,存储器的读和写信号也不能同时出现。通常,将在一个微周期中可以同时出现的微命令称为兼容性微命令,将在一个微周期内不能同时出现的微命令称为互斥性微命令。

字段直接编码是将微指令的控制字段分为若干小段,每个小段分别编码,互斥性的微命令分在同一字段内,兼容性的微命令分在不同字段内。前者可提高信息位的利用率,缩短微指令字长;后者有利于实现并行操作,加快指令的执行速度。直接编码法得到了广泛的应用,如 IBM 370、VAX-11 等都采用此编码法。

字段间接编码法是在字段直接编码法的基础上进一步缩短微指令字长的一种方法。在字段间接编码法中,一个字段的微命令编码要兼由另一字段的编码或某个标志位加以解释,以便用较少的信息位表示更多的微命令。例如,如果字段 A 为 2 位,字段 B 为 2 位,采用字段直接编码法可最多产生 8 种微命令($2^2 + 2^2$),而采用字段间接编码方法,则最多可表示 16 种微命令。不过一般每个小段都要留出一个状态,表示本字段不发出任何微命令(比如既不读也不写),所以实际上,无论字段直接编址还是字段间接编址,可表示的微命令数都会少于以上最大值。

9. 微程序控制方式

要求计算机完成的任务在确定了算法以后便可编写相应的程序,最终成为机器可直接执行的机器语言程序,而其中的任何一条机器指令可由一段微程序来解释,它们之间的关系可由图 6-6 所示。

显然,各条机器指令所对应的微程序长度可以各不相同,它取决于机器指令功能的强弱,当然也与微指令本身的功能强弱有关。于是机器指令的执行过程就成为与之相对应的微程序的执行过程,机器指令执行过程中需要的微命令由各条微指令来产生。

采用微程序控制的计算机,所有的微程序集中存放在一个独立的存储器(控制存储

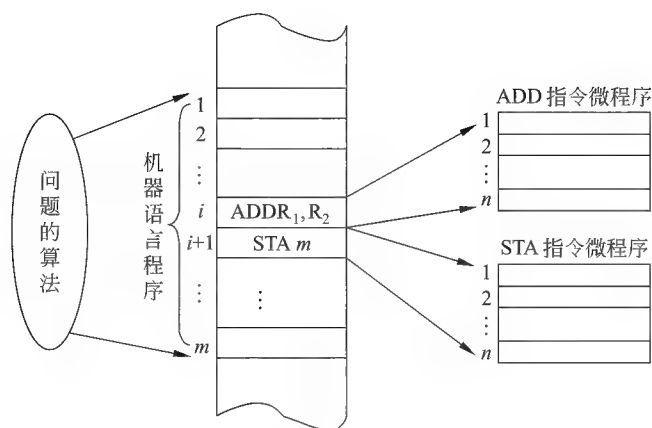


图 6-6 机器语言程序与微程序关系示意图

器)中。由于微程序一旦设计完毕,不允许改变,只允许执行,因此控存通常由 EPROM 构成,每条微指令在控存中占用一个微地址,控存的容量取决于微指令的字长和微程序的总长度。

由于一条机器指令对应一段微程序,而任何一条机器指令的取指令阶段的操作都是相同的(公操作),因此,通常将公共的部分提出来,编成一个公用微程序(取指微程序),放在控存的开始位置。这样,在机器指令对应的微程序中就只有取数、执行等阶段所需要完成的操作。当指令系统中的机器指令数为 N 种时,控制存储器中至少应当有 $N+1$ 段微程序。如果考虑将间接寻址和中断等操作也设置公用微程序,控存中微程序的数量还会更多一些。

需要提醒学生注意的是,控制存储器是 CPU 中的一部分,不要因为看见“存储器”3 个字,就将它划入存储系统的范围内。

10. 形成后继微地址的几种方式比较

形成后继微地址的方式主要有增量方式和断定方式,还有将增量方式和断定方式合二为一的结合方式。

(1) 增量方式

增量方式又称计数器法,它与用程序计数器(PC)产生下条机器指令地址的方式相类似,也有顺序执行和非顺序执行之分,因此,在微程序控制器中应当有一个微程序计数器(μPC)。在顺序执行微指令时,后续微地址由现行微地址加上一个增量来产生;在非顺序执行微指令时,由转移微指令实行转移,转移微指令的控制字段分成两部分:转移控制字段与转移地址字段。由这两个字段结合,当转移条件满足时,将转移地址字段作为下一个微地址,若转移条件不满足,则直接根据微程序计数器的内容取出下一条微指令。

增量方式的优点是简单,易于掌握,编制微程序容易,每条机器指令所对应的一段微程序一般安排在 CM 的连续单元中。增量方式的缺点是微程序中会出现大量的转移微指令,它们约占整个微指令数的 25%,导致执行时间大大增加;另外,因为要区分普通微指令和转移微指令,使得微程序控制电路复杂化。

(2) 断定方式

断定方式又称下址字段法,在微程序控制器中不需要设置微程序计数器(μPC),而是在微指令格式中设置一个下址字段,用于指明下一条要执行的微指令地址。当一条微指令被取出时,下一条微指令的地址就已获得,这相当于每条微指令都具有转移微指令的功能。

断定方式的优点是不必设置专门的转移微指令,且没有普通微指令和转移微指令的区别;但每条微指令相对增量方式中的普通微指令来说字长都比较长。

(3) 结合方式

这种方式是增量方式与断定方式的结合,此时既需要在微程序控制器中设置微程序计数器,又需要在每条微指令中都设置一个顺序控制字段,它分为两部分:转移控制字段与转移地址字段。由这两个字段结合,当转移条件满足时,将转移地址字段作为下一个微地址,若无转移要求,则直接根据微程序计数器的内容取出下一条微指令。

后继微地址的形成是一个相对比较复杂的工作,要讲清这部分内容需要花费许多时间,从目前的教学时数来看,显然不允许在这上面多下功夫。建议一般只介绍形成后继微地址几种方式的特点,不再深入讨论,如确实对这部分内容感兴趣,可以参考本章相关知识介绍中不同方式的实例。

6.3 相关知识介绍

1. 外频与前端总线频率的区别与联系

前面已经介绍过主频、外频和前端总线频率等概念。说到了外频,就不能不说前端总线频率,前端总线负责将 CPU 连接到主存,前端总线频率则直接影响 CPU 与主存交换数据的速度。过去很长一段时间里(主要是在 Pentium 4 出现之前和刚出现 Pentium 4 时),前端总线频率与外频是相同的,因此往往直接称前端总线频率为外频,随着计算机技术的发展,人们发现前端总线频率需要高于外频。

外频是 CPU 与主板之间同步运行的速度,这个概念是建立在数字脉冲信号震荡速度基础之上的;而前端总线频率是数据传输的实际速度,数据传输最大带宽取决于同时传输的数据的宽度和传输频率,即数据带宽=(总线频率 \times 数据位宽) $\div 8$ 。100MHz 外频特指数字脉冲信号在每秒钟震荡一千万次;而 100MHz 前端总线指的是每秒钟 CPU 可接受的数据传输量是 $100\text{MHz} \times 64\text{bit} \div 8\text{bit/Byte} = 800\text{MB/s}$ 。

前端总线是 CPU 和外界交换数据的最主要通道,因此前端总线的数据传输能力对计算机整体性能影响很大,如果没有足够快的前端总线,再强的 CPU 也不能明显提高计算机整体速度。目前 PC 机上所能达到的前端总线频率有 266MHz、333MHz、400MHz、533MHz、800MHz 几种,前端总线频率越高,代表着 CPU 与外界的数据传输能力越强,更能充分发挥出 CPU 的功能。较低的前端总线频率将无法保障足够的数据提供给 CPU,这样就限制了 CPU 性能的发挥,使之成为系统瓶颈。

在把外频和前端总线频率的含义清晰的分开之后,采取在脉冲信号上、下沿都进行数据传输的技术,可以使 100MHz 外频的前端总线频率为 200MHz,其实际数据传输率可以达到 1600MB/s。至此,前端总线的意义就有了进一步的飞跃,若采用了 QDR (quad data rate) 技术,或者其他类似的技术,可以使得前端总线的频率成为外频的 2 倍、4 倍甚至更高。

2. CPU 的性能

程序执行的 CPU 时间为

$$\text{CPU 时间} = \frac{\text{总时钟周期数}}{\text{时钟频率}}$$

若将程序执行过程中所处理的指令数,记为 IC。这样可以获得一个与计算机系统结构有关的参数,即“指令时钟数 CPI”。

$$\text{CPI} = \frac{\text{总时钟周期数}}{\text{IC}}$$

所以,程序执行的 CPU 时间就可以写成

$$\text{CPU 时间} = \frac{\text{CPI} \times \text{IC}}{\text{时钟频率}}$$

这个公式通常称为 CPU 性能公式。它的 3 个参数反映了与系统结构相关的下述 3 种技术。

- ① 时钟频率:反映了计算机实现技术、生产工艺和计算机组织。
- ② CPI:反映了计算机实现技术、计算机指令系统的结构和组织。
- ③ IC:反映了计算机指令级的结构和编译技术。

通过改进计算机系统设计,可以相应提高 3 个参数的指标,从而提高计算机系统的性能。从目前情况来看,提高某一个参数的性能,不会明显地影响其他两个指标。这对于综合运用各种技术改进计算机系统的性能是非常有益的。

假设计算机系统有 n 种指令,其中第 i 种指令的处理时间为 CPI_i ,在程序中第 i 种指令出现的次数为 IC_i ,则程序执行时间为

$$\text{CPU 时间} = \frac{\sum_{i=1}^n (\text{CPI}_i \times \text{IC}_i)}{\text{时钟频率}}$$

将上面两个表示 CPU 时间的公式合并,可得到

$$CPI = \frac{\sum_{i=1}^n (CPI_i \times IC_i)}{IC} = \sum_{i=1}^n CPI_i \times \frac{IC_i}{IC}$$

其中, $\frac{IC_i}{IC}$ 反映了第 i 种指令在程序中所占的比例。

例 6-5 假定在设计机器的指令系统时,对条件转移指令的设计有以下两种不同的选择:

① CPU_A 采用一条比较指令来设置相应的条件码,然后测试条件码进行转移。

② CPU_B 在转移指令中包含比较过程。

在两种 CPU 中,条件转移指令需要 2 个时钟周期,而其他的指令只需要 1 个时钟周期。又假设在 CPU_A 上,要执行的指令中有 20% 是条件转移指令,由于每条条件指令都需要一条比较指令,因此,比较指令也占用 20%。由于 CPU_A 在转移时不需要比较,因此假设它的时钟周期时间比 CPU_B 快 1.25 倍。问:哪一个 CPU 更快? 如果 CPU_A 的时钟周期时间仅仅比 CPU_B 快 1.1 倍,哪个 CPU 更快?

解: 占用 2 个时钟周期的条件转移指令占总指令的 20%,剩下的指令占用 1 个时钟周期。所以

$$CPI_A = 0.2 \times 2 + 0.8 \times 1 = 1.2$$

则总 CPU_A 时间

$$T_{CPUA} = IC_A \times CPI_A \times \text{时钟周期}_A = IC_A \times 1.2 \times \text{时钟周期}_A$$

根据假设有

$$\text{时钟周期}_B = 1.25 \times \text{时钟周期}_A$$

在 CPU_B 中没有独立的比较指令,所以 CPU_B 的程序量为 CPU_A 的 80%,转移指令的比例为

$$20\% \div 80\% = 25\%$$

这些转移指令占用 2 个时钟周期,而其余的 75% 指令只占用 1 个时钟周期,因此有

$$CPI_B = 0.25 \times 2 + 0.75 \times 1 = 1.25$$

由于 CPU_B 中没有比较指令,故

$$IC_B = 0.8 \times IC_A$$

则总 CPU_B 时间

$$T_{CPUB} = IC_B \times CPI_B \times \text{时钟周期}_B$$

$$= 0.8 \times IC_A \times 1.25 \times 1.25 \times \text{时钟周期}_A = 1.25 \times IC_A \times \text{时钟周期}_A$$

在这些假设之下,尽管 CPU_B 执行指令条数较少,但因为 CPU_A 有着更短的时钟周期,所以比 CPU_B 快。

如果 CPU_A 的时钟周期时间仅仅比 CPU_B 快 1.1 倍,则

$$\text{时钟周期}_B = 1.1 \times \text{时钟周期}_A$$

则总 CPU_B 时间

$$T_{\text{CPU}_B} = \text{IC}_B \times \text{CPI}_B \times \text{时钟周期}_B$$

$$= 0.8 \times \text{IC}_A \times 1.25 \times 1.1 \times \text{时钟周期}_A = 1.1 \times \text{IC}_A \times \text{时钟周期}_A$$

因此, CPU_B 由于执行更少的指令条数, 比 CPU_A 运行更快。

3. 微程序控制器结构

根据微程序控制器的两种不同的顺序控制方式, 微程序控制器有两种不同的结构。

增量方式要求组成一个微程序的多条微指令在控制存储器中连续存放, 微指令本身可不包含下条微指令在控制存储器中的地址。采用增量方式的微程序控制器结构框图如图 6-7 所示。

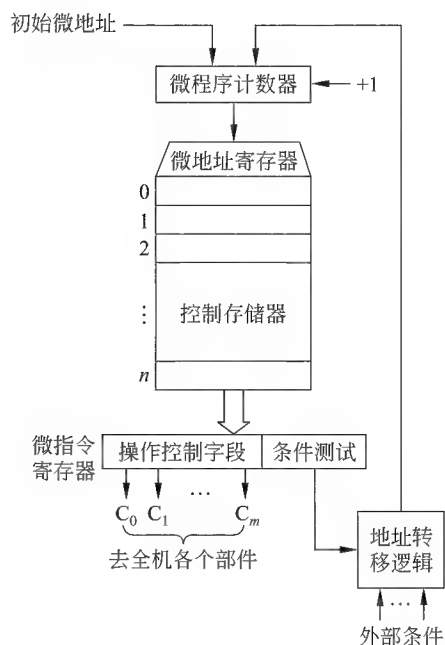


图 6-7 增量方式的微程序控制器结构框图

增量方式需要有一个微程序计数器 μPC , 一般情况下, 由 $\mu\text{PC}+1$ 来指向下条微指令在控存中的地址, 只有遇到转移类微指令才会改变 μPC 的内容以实现微程序的转移。这种结构的优点是微指令的字长有效缩短, 从而可减少控制存储器的容量。

采用断定方式的微程序控制器结构框图如图 6-8 所示。

组成各个微程序的微指令在控制存储器中可任意存放, 由各条微指令中的地址字段给出下条微指令在控存中的地址, 只有遇到条件转移类的微指令, 才需要由条件测试字段

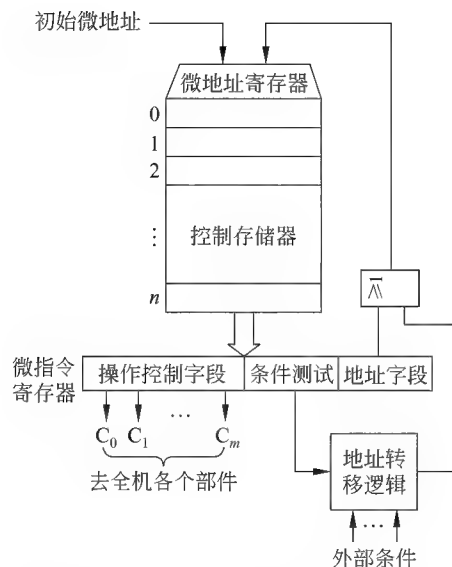


图 6 8 断定方式的微程序控制器结构框图

判定外部条件是否满足,若条件满足则地址转移逻辑修改微指令中的地址字段,以实现微程序转移的目的;若条件不满足,则按照地址字段给定的地址去执行下条微指令。

4. 后继微地址形成实例

假设某个微程序控制的计算机,其中 ADD、SUB、JC 指令的微程序流程如图 6-9 所示,为简单起见,将微命令用字符 A、B、C...代替。

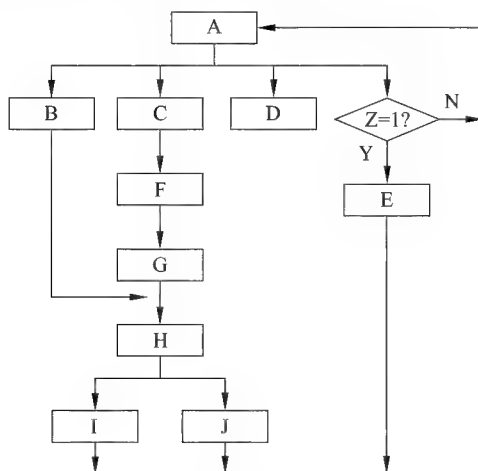


图 6 9 微程序流程图

例 6 6 用增量方式为图 6-9 表示的部分微指令序列安排微地址。由于共有 10 条微指令,再加上一些转移微指令,所以微地址至少需 5 位(用二进制表示)。

解:普通微指令中只有微命令字段,转移微指令中包括转移控制字段和转移地址字段。为了区别普通微指令和转移微指令,特增加一位标志位 T。当 T=0 时,表示此微指令为普通微指令,当 T=1 时,表示此微指令为转移微指令。两种微指令的格式如图 6-10 所示。

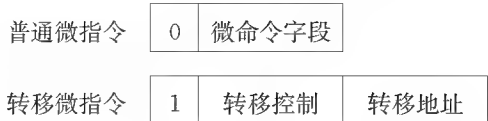


图 6 10 增量方式的微指令格式

转移地址字段的长度等于微地址寄存器的长度,转移控制字段的长度与流程图中的转移情况有关,现有 3 种转移情况,加上无条件转移需用 2 位 P1P0 来控制。

P1P0=00;无条件转移;

P1P0=01;由指令的操作码控制修改 μMAR_4 和 μMAR_3 ;

P1P0=10;由 ADD 控制修改 μMAR_0 或由 SUB 控制修改 μMAR_4 。

P1P0=11;若 Z=0,则转 00000 单元,Z=1,则 $\mu\text{PC}+1$ 。

增量方式的微地址安排如图 6-11 所示。

第一条微指令安排在 00000 单元,00000 号单元执行完后 $\mu\text{PC}+1$ 到 00001 单元,所以在 00001 单元中放一条转移微指令(T=1),实现多路分支转移,按修改方案分别转到 00010、01010、10010、11010 这 4 个单元。

① 00010 号单元执行完后 $\mu\text{PC}+1$ 到 00011 单元,所以在 00011 单元中放一条无条件转移微指令,转到 01101 单元;

② 01010 号单元执行完后 $\mu\text{PC}+1$ 到 01011 单元;01011 号单元执行完后 $\mu\text{PC}+1$ 到 01100 单元;01100 号单元执行完后 $\mu\text{PC}+1$ 到 01101 单元;01101 号单元执行完后 $\mu\text{PC}+1$ 到 01110 单元;

01110 单元中放一条转移微指令。当 ADD=1 时,修改 μMAR_0 转到 00101 单元;但 SUB=1 时,修改 μMAR_4 转到 10100 单元;

00101 号单元执行完后 $\mu\text{PC}+1$ 到 00110 单元,00110 单元中放一条无条件转移微指令,转到 00000 单元,准备取下一条机器指令;

10100 号单元执行完后 $\mu\text{PC}+1$ 到 10101 单元,10101 单元中放一条无条件转移微指令,转到 00000 单元,准备取下一条机器指令;

③ 10010 号单元执行完后 $\mu\text{PC}+1$ 到 10011 单元,所以在 10011 单元中放一条无条件转移微指令,转到 01011 单元;

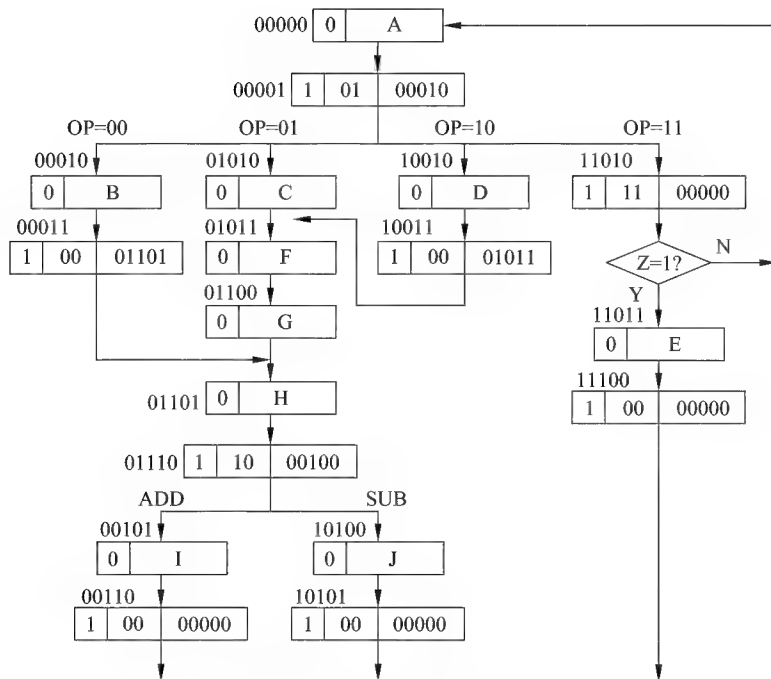


图 6-11 增量方式的微地址安排

④ 11010 单元是一条转移微指令,当 $Z=0$ 时转 00000 单元,否则 $\mu PC+1$ 到 11011 单元;

11011 号单元执行完后 $\mu PC+1$ 到 11100 单元,11100 单元中放一条无条件转移微指令,转到 00000 单元,准备取下一条机器指令。

例 6-7 用断定方式为图 6-9 表示的部分微指令序列安排微地址。

微指令中设置一个下址字段,用于指明下一条要执行的微指令地址。当一条微指令被取出时,下一条微指令的地址已获得,它相当于每条微指令都具有转移微指令的功能。

因为共有 10 条微指令,微地址需 4 位即可(用二进制表示),但在每一条微指令中均要加一个控制转移字段和一个下址字段,断定方式的微指令格式如图 6-12 所示。

微命令字段	转移控制	下址字段
-------	------	------

图 6-12 断定方式的微指令格式

下面是转移控制位 P1P0 的含义。

P1P0=00:顺序控制;

P1P0=01:由指令的操作码控制修改 μMAR_3 和 μMAR_2 ;

P1P0=10:由 ADD 控制修改 μMAR_1 或由 SUB 控制修改 μMAR_3 。

P1P0=11:由 Z 控制修改 μMAR_1 。

断定方式的微地址安排如图 6-13 所示。

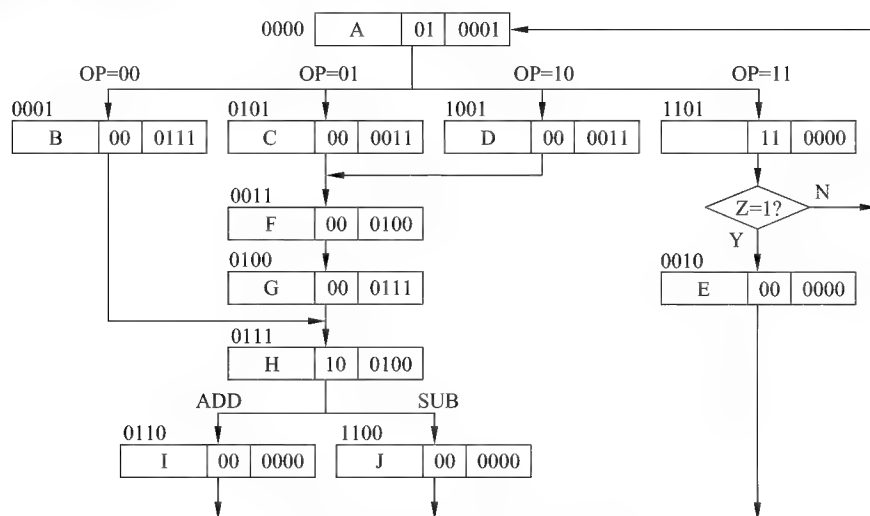


图 6-13 断定方式的微地址安排

第一条微指令安排在 0000 单元,0000 号单元是多路分支转移,按修改方案分别转到 0001、0101、1001、1101 这 4 个单元。

① 0001 号单元执行完后按顺序控制转移到 0111 单元;

② 0101 号单元执行完后按顺序控制转移到 0011 单元;0011 号单元执行完后按顺序控制转移到 0100 单元;0100 号单元执行完后按顺序控制转移到 0111 单元;

0111 号单元执行完后由 ADD 控制转移到 0110(μMAR_1)或 SUB 控制转移到 1100(μMAR_3);

0110 号单元执行完后按顺序控制转移到 0000,准备取下一条机器指令;

1100 号单元执行完后按顺序控制转移到 0000,准备取下一条机器指令;

③ 1001 号单元执行完后按顺序控制转移到 0011 单元;

④ 1101 单元是一条空操作微指令,由 Z 控制修改 μMAR_1 ,当 $Z=0$ 时转移到 0000 单元,当 $Z=1$ 时转移到 0010 单元;

0010 号单元执行完后按顺序控制转移到 0000,准备取下一条机器指令。

例 6.8 用增量方式和断定方式结合法为图 6.9 表示的部分微指令序列安排微地址。

在这种控制方式中,微指令中仍需要设置一个顺序控制字段,它分成两部分:转移控制字段与转移地址字段。由这两个字段结合,当转移条件满足时,将转移地址作为下一个

微地址,若无转移要求,这直接根据微程序计数器的内容取出下一条微指令。

因为共有 10 条微指令,微地址需 4 位即可(用二进制表示),有 3 种转移情况,考虑计数控制和无条件转移,需用 3 位 P2、P1、P0 来控制。增量和断定结合方式的微指令的格式如图 6-14 所示。

微命令字段	转移控制字段 BCF	转移地址字段 BAF
-------	------------	------------

图 6 14 增量和断定结合方式的微指令格式

P2P1P0=000;由 μPC 计数得到下一微地址;
P2P1P0=001;无条件转移;
P2P1P0=010;由 ADD 控制修改 μMAR_2 或由 SUB 控制修改 μMAR_1 ;
P2P1P0=011;由指令的操作码控制修改 μMAR_3 和 μMAR_2 ;
P2P1P0=100;由 Z 控制。
增量和断定结合方式的微地址安排如图 6-15 所示。

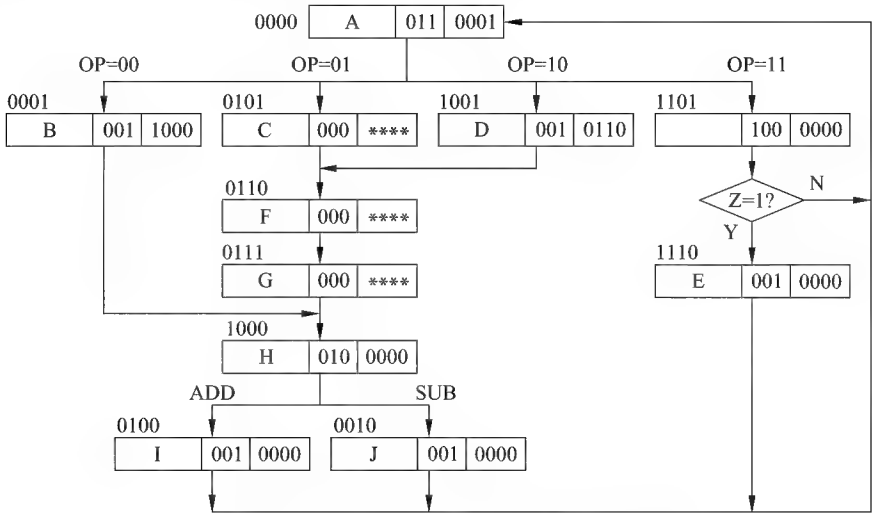


图 6-15 增量和断定结合方式的微地址安排

第一条微指令安排在 0000 单元,0000 号单元是多路分支转移,按修改方案分别转到 0001、0101、1001、1101 这 4 个单元。

- ① 0001 号单元执行完后无条件转移到 1000;
- ② 0101 号单元执行完后计数到地址 0110;0110 号单元执行完后计数到地址 0111;0111 号单元执行完后计数到地址 1000;
- 1000 号单元执行完后由 ADD 控制转移到 0100 (μMAR_2)或由 SUB 控制转移到

0010(μMAR_1);

0100 号单元执行完后无条件转移到 0000,准备取下一条机器指令;

0010 号单元执行完后无条件转移到 0000,准备取下一条机器指令;

③ 1001 号单元执行完后无条件转移到 0110;

④ 1101 单元是一条空操作微指令,当 $Z=0$ 时转 0000 单元,否则 $\mu\text{PC}+1$ 到 1110 单元;

1110 号单元执行完后无条件转移到 00000 单元,准备取下一条机器指令。

5. 微程序设计举例

例 6-9 设某计算机的 CPU 结构如图 6-16 所示。A、B、C 均为 8 位寄存器,它们的输入和输出的控制信号分别为 IN_A , IN_B , IN_C 和 OUT_A , OUT_B , OUT_C ; A、C 还可以级联右移,其移位控制信号为 SHT_{AC} ; A 的清空控制信号为 CLR_A ; D 为计数器,其置数控制信号为 SET_D ,减 1 计数器控制信号为 DEC_D ; Z 和 S 为状态信号,当 $D=0$ 时, $Z=1$, S 为 C 寄存器的最低位; + 为 ALU 的加法控制信号。

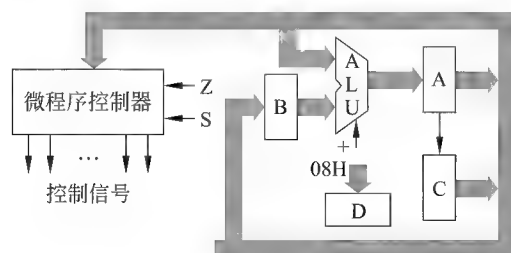


图 6-16 某机的 CPU 结构

该计算机采用微程序控制,微指令格式如图 6-17 所示。



图 6-17 某机的微指令格式

图 6-17 中,CF 为控制字段,采用直接控制法,控制信号共 11 位,从高位到低位的顺序为: OUT_A , OUT_B , OUT_C , IN_A , IN_B , IN_C , +, CLR_A , SET_D , DEC_D , SHT_{AC} 。

BCF 为转移控制字段,共 2 位,其含义为:

00 顺序

01 测试 Z

10 测试 S

11 取微指令

BAF 为转移地址字段,转移地址 3 位。

设 B、C 分别存放被乘数和乘数,且均为无符号定点小数。

编址实现 $B \times C \rightarrow A, C$ (A 存放高位积, C 存放低位积) 的微程序。

解: 实现 $B \times C \rightarrow A, C$ (A 存放高位积, C 存放低位积) 的操作流程如图 6-18(a) 所示, 微程序流程图如图 6-18(b) 所示。

微程序采用增量与断定结合方式, 微地址的安排见图 6-19。

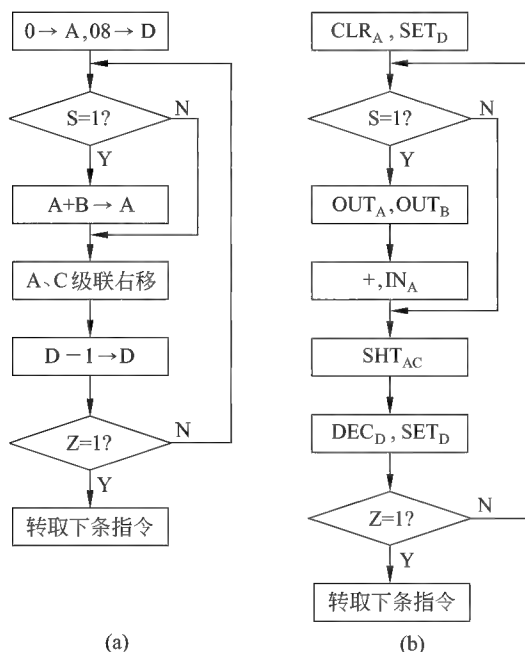


图 6-18 实现 $B \times C \rightarrow A, C$ 的操作流程与微程序流程

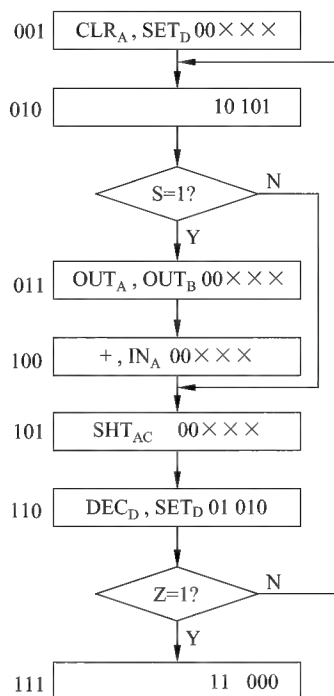


图 6-19 微地址安排

假设取指微指令放在 000 单元中, 取指后进入 001 单元, 开始执行乘法微程序。

001 单元中的微指令执行完后, 顺序执行 010 单元中的微指令;

010 单元中测试位为 10, 表示测试 S, 若 $S=0$, 则转 101 单元, 否则顺序执行 011 单元中的微指令;

011 单元中的微指令执行完后, 顺序执行 100 单元中的微指令;

100 单元中的微指令执行完后, 顺序执行 101 单元中的微指令;

101 单元中的微指令执行完后, 顺序执行 110 单元中的微指令;

110 单元中的测试位为 01, 表示测试 Z, 若 $Z=0$, 则转 010 单元, 否则顺序执行 111 单元中的微指令;

111 单元中测试位为 11, 表示转取指微指令。

综上所述,实现 $B \times C \rightarrow A, C$ (A 存放高位积, C 存放低位积) 的微程序如表 6-1 所示。

表 6-1 实现 $B \times C \rightarrow A, C$ 的微程序

微地址	OUT _A	OUT _B	OUT _C	IN _A	IN _B	IN _C	+	CLR _A	SET _D	DEC _D	SHT _{AC}	BCF	BAF
001	0	0	0	0	0	0	0	1	1	0	0	00	***
010	0	0	0	0	0	0	0	0	0	0	0	10	101
011	1	1	0	0	0	0	0	0	0	0	0	00	***
100	0	0	0	1	0	0	1	0	0	0	0	00	***
101	0	0	0	0	0	0	0	0	0	0	1	00	***
110	0	0	0	0	0	0	0	0	1	1	0	01	010
111	0	0	0	0	0	0	0	0	0	0	0	11	000

6. 毫微程序设计和毫微程序控制器

根据微指令操作控制字段的编码方法,微指令可归纳为水平型微指令和垂直型微指令两种。

操作控制字段采用直接控制法、字段编码法的微指令一般属于水平型微指令,其特点是:①微指令字较长;②微指令并行操作能力强,即在一个微周期中同时可执行多个微命令;③微指令结构与机器指令的差别大。采用水平型微指令来编写微程序,称为水平微程序设计。只有精通机器结构、数据通路、时序系统及微指令编码的专业人员才能进行这种微程序设计,一般用户难以设计。

操作控制字段采用最短编码法的微指令属于垂直型微指令,其特点是:①微指令字较短;②微指令并行能力差,一条微指令一般只能控制数据通路的 1~2 种信息传送;③微指令结构与机器指令相似。采用垂直型微指令来编写微程序,称为垂直微程序设计。这种微程序设计只需了解微指令的功能,而对数据通路结构可不必考虑,所以这种微程序便于用户编写。

如果把垂直微程序设计和水平微程序设计结合起来,采用两级微程序设计方法就称为毫微程序设计。第一级为垂直微程序,用来解释机器指令,故仍可称之为微程序,存放在微程序的控制存储器(μCM)中。第二级为水平微程序,用来解释垂直型微指令,并产生相应微命令,实现对数据通路的控制。由于此时的水平微程序是解释微程序的微程序,所以被称为毫微程序,存放在毫微程序的控制存储器(nCM)中。所以可以说毫微程序设计就是用水平型毫微指令来解释垂直型微指令的微程序设计。

垂直型微程序是根据指令系统和有关处理过程的需要来编制的,它有严格的顺序结构。由于垂直型微指令很像机器指令,编程过程就像机器指令编程一样。水平型微指令(毫微指令)是由第一级调用的,它具有并行操作控制能力,但不包含后继微地址信息,各

条毫微指令之间没有顺序关系。若干条垂直型微指令可以调用同一条毫微指令,所以在第二级控存中的每条毫微指令都是不相同的。

毫微程序控制器的如图 6-20 所示,它与通常的微程序控制器相比,除了增加了存放毫微程序的控制存储器 nCM 外,还增加了毫微地址寄存器 $nMAR$ 和毫微指令寄存器 nIR 及其译码电路。在毫微程序的具体设计中,可能有以下 3 种情况:

- ① 微命令是极简单的控制信号,可由垂直微指令直接产生,这就无需再用毫微指令去解释;
- ② 一条垂直型微指令只用一条毫微指令来解释;
- ③ 一条垂直型微指令由一段毫微程序来解释,此时毫微程序与垂直型微指令的关系就相当于微程序与机器指令的关系。

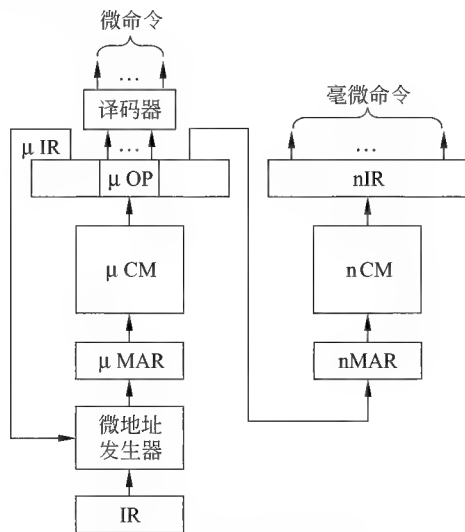


图 6 20 毫微程序控制器结构

当从 μCM 中读出一条微指令,微操作码字段经译码后可以产生一些简单的微命令,同时还给出一个对应的毫微地址,以便需要时可从 nCM 中取出一条毫微指令,用若干微命令来解释该微指令的操作,以实现数据通路和其他处理过程的控制。

毫微程序设计的主要优点有:

- ① 可以减少控制存储器的总容量, μCM 的横向容量很短, nCM 的纵向容量很小;
- ② 用垂直型微指令编制微程序比较容易;
- ③ 效率高,可充分利用数据通路;
- ④ 独立性强,毫微程序之间没有顺序关系,任意修改、增删毫微指令都不会影响毫微程序的控制结构;

⑤ 灵活性好,若想改变机器指令的功能,只需修改垂直型微程序,无需改变毫微程序,因此能方便地修改和扩充指令的功能,具有动态结构的特点。

毫微程序设计的缺点是不易做到高速度,一个微周期内可能要执行一条微指令和一条毫微指令,两次访问控制存储器,速度将受到影响。另外,也增加了硬件的复杂性和成本,所以小型机、微型机一般不采用。

7. 流水线的性能分析

流水线是把复杂的过程分解为若干个子过程,每个子过程由一个独立的功能部件来完成,处理对象在各子过程连成的线路上连续流动。在同一时间内完成对不同子过程的处理。衡量流水线性能的主要指标有吞吐率、加速比和效率。

(1) 吞吐率

吞吐率 TP 是指在单位时间内流水线所完成的任务数或输出的结果数。

如果流水线各段的经过时间相同,流水线的最大吞吐率 $TP_{\max} = \frac{1}{\Delta t}$ 。如果流水线各段的经过时间不同,流水线的最大吞吐率 $TP_{\max} = \frac{1}{\max\{\Delta t_1, \dots, \Delta t_i, \dots, \Delta t_n\}}$,此时受限于流水线中最慢子过程经过的时间。流水线中经过时间最长的子过程称为瓶颈子过程。

由于流水开始时总要有有一段建立时间、结束时又需要有排空的时间、多功能流水时某些段可能闲置未用、功能切换时流水线也需要排空、重组等诸多原因,流水线的实际吞吐率 TP 一般显著地低于最大吞吐率 TP_{\max} 。设 m 段流水线的各段经过时间均为 Δt_0 ,则需要 $T_0 = m\Delta t_0$ 的流水建立时间,之后每隔 Δt_0 就可流出一条指令,完成 n 个任务的解释共需时间 $T = m\Delta t_0 + (n-1)\Delta t_0$,流水线的实际吞吐率为

$$TP = \frac{n}{m\Delta t_0 + (n-1)\Delta t_0} = \frac{1}{\Delta t_0 \left(1 + \frac{m-1}{n}\right)} = \frac{TP_{\max}}{1 + \frac{m-1}{n}}$$

(2) 加速比

加速比 S_p 是指完成同样一批任务,不使用流水线所用的时间与使用流水线所用时间之比。

$$S_p = \frac{n \cdot m \cdot \Delta t_0}{m\Delta t_0 + (n-1)\Delta t_0} = \frac{m}{1 + \frac{m-1}{n}}$$

(3) 效率

效率 η 是指流水线中设备的实际使用时间占整个运行时间之比。

由于流水线存在有建立时间和排空时间,在连续完成 n 个任务的时间里,各段并不总是满负荷工作。

如果是线性流水线且各段经过的时间相同,流水线的效率正比于吞吐率,即:

$$\eta = \frac{n}{n + (m - 1)} = TP \cdot \Delta t_0$$

对于非线性流水或线性流水但各段经过的时间不等时,上式的关系就不存在了,只有通过画实际工作的时空图才能求出吞吐率和效率。整个流水线的效率为:

$$\eta = \frac{n \text{ 个任务实际占用的时空图}}{m \text{ 个段总的时空图}}$$

8. 消除流水线瓶颈的方法

为了提高流水线的最大吞吐率,首先要找出瓶颈,然后设法消除此瓶颈。例如,有一个4段的指令流水线如图6-21(a)所示,其中1、3、4段的经过时间均为 Δt_0 ,只有第2段的经过时间为 $3\Delta t_0$,因此瓶颈段在2段,使整个流水线最大吞吐率只有 $\frac{1}{3\Delta t_0}$ 。

消除瓶颈的一种方法是将瓶颈子过程再细分,例如,将2段再细分成21、22、23这3个子段,如图6-21(b)所示。让各子段经过时间都减少到 Δt_0 ,这样,最大吞吐率就可提高到 $\frac{1}{\Delta t_0}$ 。

消除瓶颈的另一种方法是瓶颈子过程并联。假设2段已不能再细分,则可以通过重复设置3套瓶颈段(2a、2b、2c)并联,让它们交叉并行,如图6-21(c)所示。每隔 Δt_0 轮流给其中一个瓶颈段分配任务,使最大吞吐率提高到 $\frac{1}{\Delta t_0}$ 。这种方法比瓶颈子过程再细分控制要复杂,设备量要多。

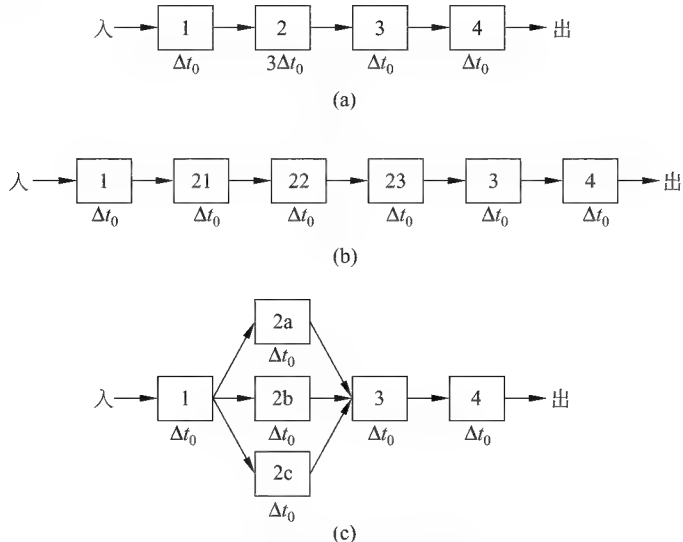


图 6 21 流水线的瓶颈及消除瓶颈的方法

如果线性流水线每段经过时间 Δt_i 不等,其中瓶颈段的时间为 Δt_j ,则完成 n 个任务所能达到的

$$\text{实际吞吐率 } TP = \frac{n}{\sum_{i=1}^m \Delta t_i + (n-1)\Delta t_j}$$

$$\text{加速比 } S_P = \frac{n \cdot \sum_{i=1}^m \Delta t_i}{\sum_{i=1}^m \Delta t_i + (n-1)\Delta t_j}$$

$$\text{效率 } \eta = \frac{n \cdot \sum_{i=1}^m \Delta t_i}{m \cdot \left[\sum_{i=1}^m \Delta t_i + (n-1)\Delta t_j \right]}$$

6.4 教材习题解答

1. 控制器有哪几种控制方式? 各有何特点?

解: 控制器的控制方式可以分为 3 种:同步控制方式、异步控制方式和联合控制方式。

同步控制方式的各项操作都由统一的时序信号控制,在每个机器周期中产生统一数目的节拍电位和工作脉冲。这种控制方式设计简单,容易实现;但是对于许多简单指令来说会有较多的空闲时间,造成较大数量的时间浪费,从而影响了指令的执行速度。

异步控制方式的各项操作不采用统一的时序信号控制,而根据指令或部件的具体情况决定,需要多少时间,就占用多少时间。异步控制方式没有时间上的浪费,因而提高了机器的效率,但是控制比较复杂。

联合控制方式是同步控制和异步控制相结合的方式。

2. 什么是一级时序系统?

解: 三级时序系统是指机器周期、节拍和工作脉冲。计算机中每个指令周期划分为若干个机器周期,每个机器周期划分为若干个节拍,每个节拍中设置一个或几个工作脉冲。

3. 控制器有哪些基本功能? 它可分为哪几类? 分类的依据是什么?

解: 控制器的基本功能有:

- (1) 从主存中取出一条指令,并指出下一条指令在主存中的位置。
- (2) 对指令进行译码或测试,产生相应的操作控制信号,以便启动规定的动作。

(3) 指挥并控制 CPU、主存和输入输出设备之间的数据流动。

控制器可分为组合逻辑型、存储逻辑型、组合逻辑与存储逻辑结合型 3 类,分类的依据在于控制器的核心——微操作信号发生器(控制单元 CU)的实现方法不同。

4. 中央处理器有哪些功能? 它由哪些基本部件所组成?

解: 从程序运行的角度来看, CPU 的基本功能就是对指令流和数据流在时间与空间上实施正确的控制。对于冯·诺依曼结构的计算机而言,数据流是根据指令流的操作而形成的,也就是说数据流是由指令流来驱动的。

中央处理器由运算器和控制器组成。

5. 中央处理器中有哪几个主要寄存器? 试说明它们的结构和功能。

解: CPU 中的寄存器是用来暂时保存运算和控制过程中的中间结果、最终结果及控制、状态信息的,它可分为通用寄存器和专用寄存器两大类。

通用寄存器可用来存放原始数据和运算结果,有的还可以作为变址寄存器、计数器、地址指针等。专用寄存器是专门用来完成某一种特殊功能的寄存器,如程序计数器 PC、指令寄存器 IR、存储器地址寄存器 MAR、存储器数据寄存器 MDR、状态标志寄存器 PSWR 等。

6. 某机 CPU 芯片的主振频率为 8 MHz,其时钟周期是多少 μs ? 若已知每个机器周期平均包含 4 个时钟周期,该机的平均指令执行速度为 0.8 MIPS,试问:

(1) 平均指令周期是多少 μs ?

(2) 平均每个指令周期含有多少个机器周期?

(3) 若改用时钟周期为 $0.4\mu\text{s}$ 的 CPU 芯片,则计算机的平均指令执行速度又是多少 MIPS?

(4) 若要得到 40 万次/s 的指令执行速度,则应采用主振频率为多少 MHz 的 CPU 芯片?

解: 时钟周期 $= 1 \div 8 \text{ MHz} = 0.125\mu\text{s}$

(1) 平均指令周期 $= 1 \div 0.8 \text{ MIPS} = 1.25\mu\text{s}$

(2) 机器周期 $= 0.125\mu\text{s} \times 4 = 0.5\mu\text{s}$

平均每个指令周期的机器周期数 $= 1.25\mu\text{s} \div 0.5\mu\text{s} \div 4 = 2.5$

(3) 平均指令执行速度 $= \frac{1}{0.4 \times 4 \times 2.5} = 0.25 \text{ MIPS}$

(4) 主振频率 $= 4 \text{ MHz}$

7. 以一条典型的单地址指令为例,简要说明下列部件在计算机的取指周期和执行周期中的作用。

(1) 程序计数器 PC;

(2) 指令寄存器 IR;

(3) 算术逻辑运算部件 ALU ;

(4) 存储器数据寄存器 MDR ;

(5) 存储器地址寄存器 MAR。

解：(1) 程序计数器 PC :存放指令地址；

(2) 指令寄存器 IR :存放当前指令；

(3) 算术逻辑运算部件 ALU :进行算逻运算；

(4) 存储器数据寄存器 MDR :存放写入或读出的数据/指令；

(5) 存储器地址寄存器 MAR :存放写入或读出的数据/指令的地址。

以单地址指令“加 1 (INC A)”为例,该指令分为 3 个周期:取指周期、分析取数周期、执行周期。3 个周期完成的操作如表 6-2 所示。

表 6 2 加 1 指令完成的操作

	取指周期	分析取数周期	执行周期
PC	(PC)→MAR		
IR	指令→MDR→IR		
ALU	(PC)+1		(A)+1
MAR	指令地址→MAR	A→MAR	
MDR	指令→MDR	(A)→MDR	(A)+1→MDR

8. 什么是指令周期? 什么是 CPU 周期? 它们之间有什么关系?

解：指令周期是指取指令、分析取数到执行指令所需的全部时间。CPU 周期(机器周期)是完成一个基本操作的时间。一个指令周期划分为若干个 CPU 周期。

9. 指令和数据都存放在主存,如何识别从主存储器中取出的是指令还是数据?

解：指令和数据都存放在主存,它们都以二进制代码形式出现,区分的方法为:

(1) 取指令或数据时所处的机器周期不同:取指周期取出的是指令;分析取数或执行周期取出的是数据。

(2) 取指令或数据时地址的来源不同:指令地址来源于程序计数器;数据地址来源于地址形成部件。

10. CPU 中指令寄存器是否可以不要? 指令译码器是否能直接对存储器数据寄存器 MDR 中的信息译码? 为什么? 请以无条件转移指令 JMP A 为例说明。

解：指令寄存器不可以不要。指令译码器不能直接对 MDR 中的信息译码,因为在取指周期 MDR 的内容是指令,而在取数周期 MDR 的内容是操作数。以 JMP A 指令为例,假设指令占两个字,第一个字为操作码,第二个字为转移地址,它们从主存中取出时都需要经过 MDR,其中只有第一个字需要送至指令寄存器,并且进行指令的译码,而第二个字不需要送指令寄存器。

11. 设一地址指令格式如下:

@	OP	A
---	----	---

现在有 4 条一地址指令:LOAD(取数)、ISZ(加“1”为零跳)、DSZ(减“1”为零跳)、STORE(存数),在一台单总线单累加器结构的机器上运行,试排出这 4 条指令的微操作序列。要求:当排 ISZ 和 DSZ 指令时不要破坏累加寄存器 Acc 原来的内容。

解:(1) LOAD(取数)指令

PC→MAR,READ ;取指令
MM→MDR
MDR→IR,PC+1→PC
A→MAR,READ ;取数据送 Acc
MM→MDR
MDR→Acc

(2) ISZ(加“1”为零跳)指令

取指令微操作略。

A→MAR,READ ;取数据送 Acc
MM→MDR
MDR→Acc
Acc+1→Acc ;加 1
If Z=1 then PC+1→PC ;结果为 0,PC+1
Acc→MDR,WRITE ;保存结果
MDR→MM
Acc-1→Acc ;恢复 Acc

(3) DSZ(减“1”为零跳)指令

取指令微操作略。

A→MAR,READ ;取数据送 Acc
MM→MDR
MDR→Acc
Acc-1→Acc ;减 1
If Z=1 then PC+1→PC ;结果为 0,PC+1
Acc→MDR,WRITE ;保存结果
MDR→MM
Acc+1→Acc ;恢复 Acc

(4) STORE(存数)指令:

取指令微操作略。

$A \rightarrow \text{MAR}$; Acc 中的数据写入主存单元
 $\text{Acc} \rightarrow \text{MDR}, \text{WRITE}$
 $\text{MDR} \rightarrow \text{MM}$

12. 某计算机的 CPU 内部结构如图 6-22 所示。两组总线之间的所有数据传送通过 ALU。ALU 还具有完成以下功能的能力:

$F = A$; $F = B$
 $F = A + 1$; $F = B + 1$
 $F = A - 1$; $F = B - 1$

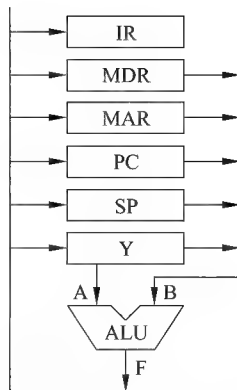


图 6-22 某机 CPU 内部结构

写出转子指令 (JSR) 的取指和执行周期的微操作序列。JSR 指令占两个字, 第一个字是操作码, 第二个字是子程序的入口地址。返回地址保存在存储器堆栈中, 堆栈指示器始终指向栈顶。

解: ① $\text{PC} \rightarrow \text{B}, F = \text{B}, F \rightarrow \text{MAR}, \text{Read}$; 取指令的第一个字
 ② $\text{PC} \rightarrow \text{B}, F = \text{B} + 1, F \rightarrow \text{PC}$
 ③ $\text{MDR} \rightarrow \text{B}, F = \text{B}, F \rightarrow \text{IR}$
 ④ $\text{PC} \rightarrow \text{B}, F = \text{B}, F \rightarrow \text{MAR}, \text{Read}$; 取指令的第二个字
 ⑤ $\text{PC} \rightarrow \text{B}, F = \text{B} + 1, F \rightarrow \text{PC}$
 ⑥ $\text{MDR} \rightarrow \text{B}, F = \text{B}, F \rightarrow \text{Y}$
 ⑦ $\text{SP} \rightarrow \text{B}, F = \text{B} - 1, F \rightarrow \text{SP}, F \rightarrow \text{MAR}$; 修改栈指针, 返回地址压入堆栈
 ⑧ $\text{PC} \rightarrow \text{B}, F = \text{B}, F \rightarrow \text{MDR}, \text{Write}$
 ⑨ $\text{Y} \rightarrow \text{A}, F = \text{A}, F \rightarrow \text{PC}$; 子程序的首地址 $\rightarrow \text{PC}$
 ⑩ End

13. 某机主要部件如图 6-23 所示。

(1) 请补充各部件间的主要连接线, 并注明数据流动方向。

(2) 拟出指令 $\text{ADD}(\text{R}_1), (\text{R}_2) +$ 的执行流程 (含取指过程与确定后继指令地址)。该指令的含义是进行加法操作, 源操作数地址和目的操作数地址分别在寄存器 R_1 和 R_2 中, 目的操作数寻址方式为自增型寄存器间址。

解: (1) 将各部件间的主要连接线补充完后如图 6-24 所示。

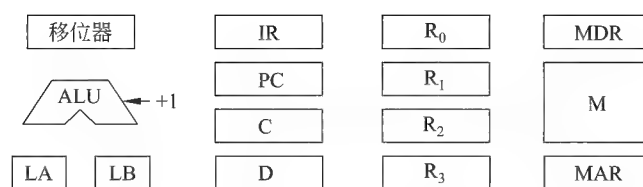
(2) 指令 $\text{ADD}(\text{R}_1), (\text{R}_2) +$ 的含义为

$((\text{R}_1)) + ((\text{R}_2)) \rightarrow (\text{R}_2)$

$(\text{R}_2) + 1 \rightarrow \text{R}_2$

指令的执行流程如下:

① $(\text{PC}) \rightarrow \text{MAR}$; 取指令



注：LA — A 输入选择器
LB — B 输入选择器
C、D — 暂存器

图 6 23 某机主要部件

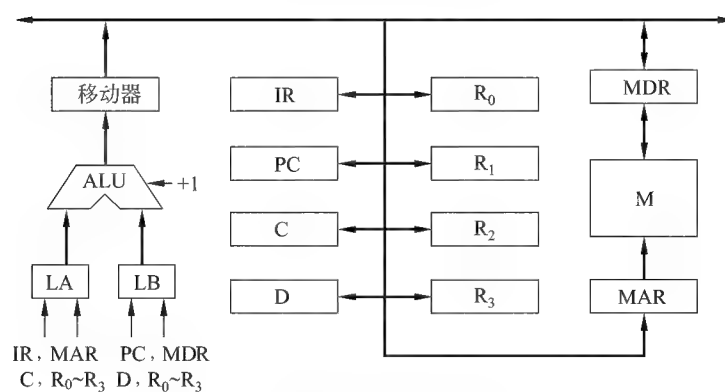


图 6 24 某机数据通路

- ② Read
- ③ $M(MAR) \rightarrow MDR \rightarrow IR$
- ④ $(PC) + 1 \rightarrow PC$
- ⑤ $(R_1) \rightarrow MAR$;取被加数
- ⑥ Read
- ⑦ $M(MAR) \rightarrow MDR \rightarrow C$
- ⑧ $(R_2) \rightarrow MAR$;取加数
- ⑨ Read
- ⑩ $M(MAR) \rightarrow MDR \rightarrow D$
- ⑪ $(R_2) + 1 \rightarrow R_2$;修改目的地址
- ⑫ $(C) + (D) \rightarrow MDR$;求和并保存结果
- ⑬ Write
- ⑭ $MDR \rightarrow MM$

14. CPU 结构如图 6-25 所示,其中有一个累加寄存器 AC、一个状态条件寄存器和其

他 4 个寄存器,各部件之间的连线表示数据通路,箭头表示信息传送方向。

- (1) 标明 4 个寄存器的名称。
- (2) 简述指令从主存取出送到控制器的数据通路。
- (3) 简述数据在运算器和主存之间进行存取访问的数据通路。

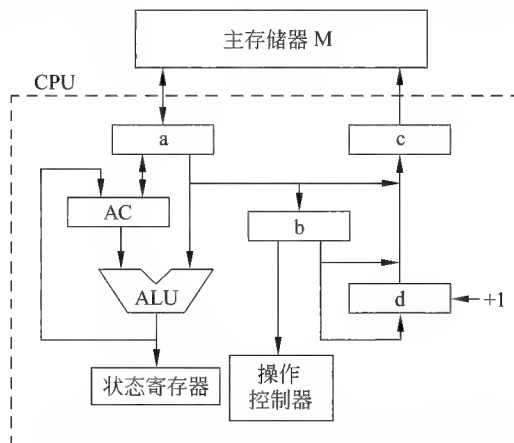


图 6 25 某机 CPU 结构

解：(1) 这 4 个寄存器中,a 为存储器数据寄存器 MDR,b 为指令寄存器 IR,c 为存储器地址寄存器 MAR,d 为程序计数器 PC。

(2) 取指令的数据通路：

$PC \rightarrow MAR \rightarrow MM \rightarrow MDR \rightarrow IR$

(3) 数据从主存中取出的数据通路(设数据地址为 X)：

$X \rightarrow MAR \rightarrow MM \rightarrow MDR \rightarrow ALU \rightarrow AC$

数据存入主存中的数据通路(设数据地址为 Y)：

$Y \rightarrow MAR, AC \rightarrow MDR \rightarrow MM$

15. 什么是微命令和微操作? 什么是微指令? 微程序和机器指令有何关系? 微程序和程序之间有何关系?

解：微命令是控制计算机各部件完成某个基本微操作的命令。微操作是指计算机中最基本的、不可再分解的操作。微命令和微操作是一一对应的,微命令是微操作的控制信号,微操作是微命令的操作过程。

微指令是若干个微命令的集合。

微程序是机器指令的实时解释器,每一条机器指令都对应一个微程序。

微程序和程序是两个不同的概念。微程序是由微指令组成的,用于描述机器指令,实际上是机器指令的实时解释器,微程序是由计算机的设计者事先编制好并存放在控制存储器中的,一般不提供给用户;程序是由机器指令组成的,由程序员事先编制好并存放在

主存储器中。

16. 什么是垂直型微指令? 什么是水平型微指令? 它们各有什么特点? 又有什么区别?

解: 垂直型微指令是指一次只能执行一个微命令的微指令; 水平型微指令是指一次能定义并能并行执行多个微命令的微指令。

垂直型微指令的并行操作能力差, 一般只能实现一个微操作, 控制 1~2 个信息传送通路, 效率低, 执行一条机器指令所需的微指令数目多, 执行时间长; 但是微指令与机器指令很相似, 所以容易掌握和利用, 编程比较简单, 不必过多地了解数据通路的细节, 且微指令字较短。水平型微指令的并行操作能力强, 效率高, 灵活性强, 执行一条机器指令所需微指令的数目少, 执行时间短; 但微指令字较长, 增加了控存的横向容量, 同时微指令和机器指令的差别很大, 设计者只有熟悉了数据通路, 才有可能编制出理想的微程序, 一般用户不易掌握。

17. 水平型和垂直型微程序设计之间各有什么区别? 串行微程序设计和并行微程序设计有什么区别?

解: 水平型微程序设计是面对微处理器内部逻辑控制的描述, 所以把这种微程序设计方法称为硬方法; 垂直型微程序设计是面向算法的描述, 所以把这种微程序设计方法称为软方法。

在串行微程序设计中, 取微指令和执行微指令是顺序进行的, 在一条微指令取出并执行之后, 才能取下一条微指令; 在并行微程序设计中, 将取微指令和执行微指令的操作重叠起来, 从而缩短微周期。

18. 图 6-26 给出了某微程序控制计算机的部分微指令序列。图中每一框代表一条微指令。分支点 a 由指令寄存器 IR 的第 5、6 两位决定。分支点 b 由条件码 C_0 决定。现

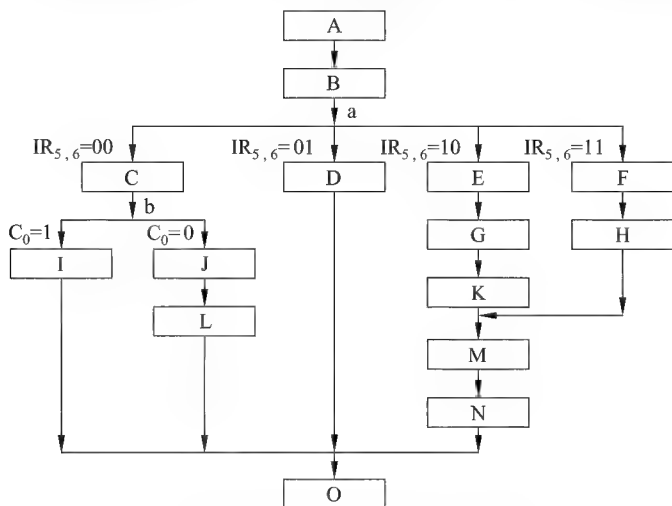


图 6-26 某机的部分微指令序列

采用下址字段实现该序列的顺序控制。已知微指令地址寄存器字长 8 位。

- (1) 设计实现该微指令序列的微指令字之顺序控制字段格式。
- (2) 给出每条微指令的二进制编码地址。
- (3) 画出微程序控制器的简化框图。

解：(1) 该微程序流程有两处有分支的地方，第一处有 4 路分支，由指令操作码 $IR_5 IR_6$ 指向 4 条不同的微指令，第二处有 2 路分支，根据运算结果 C_0 的值决定后继微地址。加上顺序控制，转移控制字段取 2 位。图 6-26 中共有 15 条微指令，则下址字段至少需要 4 位，但因已知微指令地址寄存器字长 8 位 ($\mu MAR_7 \sim \mu MAR_0$)，故下址字段取 8 位。微指令的顺序控制字段格式如图 6-27 所示。

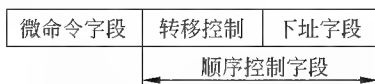


图 6 27 微指令的顺序控制字段格式

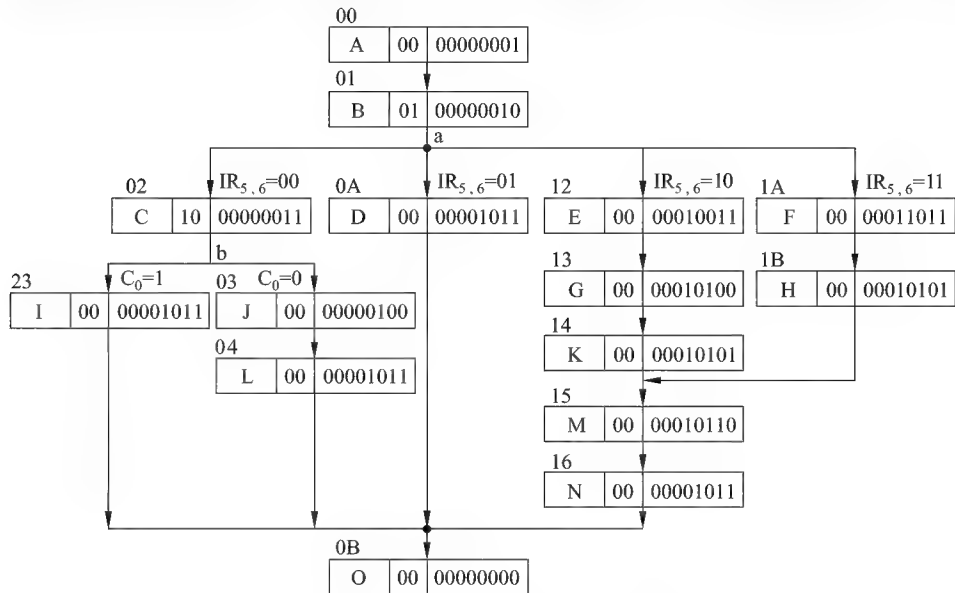
(2) 转移控制字段 2 位：

00 顺序控制

01 由 $IR_5 IR_6$ 控制修改 $\mu MAR_4, \mu MAR_3$ 。

10 由 C_0 控制修改 μMAR_5 。

微程序流程的微地址安排如图 6-28 所示。每条微指令的二进制编码地址见表 6-3。



注：每条微指令前的微地址用十六进制表示。

图 6 28 微程序流程的微地址安排

表 6-3 每条微指令的二进制编码地址

微 指 令			
微地址	操作控制字段	顺序控制字段	
二进制	微命令	测试判别	下地址
00000000	A	00	00000001
00000001	B	01	00000010
00000010	C	10	00000011
00000011	J	00	00000100
00000100	L	00	00001011
00001010	D	00	00001011
00001011	O	00	00000000
00010010	E	00	00010011
00010011	G	00	00010100
00010100	K	00	00010101
00010101	M	00	00010110
00010110	N	00	00001011
00011010	F	00	00011011
00011011	H	00	00010101
00100011	I	00	00001011

(3) 微程序控制器的简化框图略。

19. 已知某机采用微程序控制方式,其控制存储器容量 512×48 位,微程序可在整个控制存储器中实现转移,可控制转移的条件共 4 个,微指令采用水平型格式,后继指令地址采用断定方式,微指令格式如图 6-29 所示。

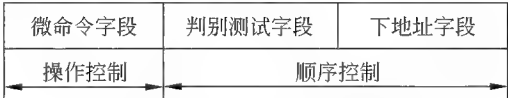


图 6-29 某机的微指令格式

- (1) 微指令中的 3 个字段分别应为多少位?
- (2) 画出围绕这种微指令格式的微程序控制器逻辑框图。

解: (1) 因为控制转移的条件共 4 个,则判别测试字段为 2 位;因为控存容量为 512 个单元,所以下地址字段为 9 位;微命令字段是 $(48-2-9)=37$ 位。

(2) 对应上述微指令格式的微程序控制器逻辑框图如图 6-30。

20. 某机有 8 条微指令 $I_1\sim I_8$, 每条微指令所含的微命令控制信号如表 6-4 所列。

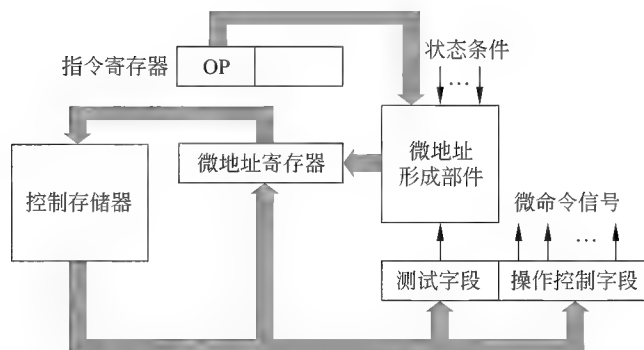


图 6 30 微程序控制器逻辑框图

表 6 4 微指令所含微命令控制信号

微指令	微命令信号									
	a	b	c	d	e	f	g	h	i	j
I ₁	✓	✓	✓	✓	✓					
I ₂	✓			✓		✓	✓			
I ₃		✓						✓		
I ₄			✓							
I ₅			✓		✓		✓		✓	
I ₆	✓							✓		✓
I ₇			✓	✓				✓		
I ₈	✓	✓						✓		

a~j 分别代表 10 种不同性质的微命令信号,假设一条微指令的操作控制字段为 8 位,请安排微指令的操作控制字段格式,并将全部微指令代码化。

解: 因为微指令的操作控制字段只有 8 位,所以不能采用直接控制法。又因为微指令中有多个微命令是兼容性的微命令,如微指令 I₁ 中的微命令 a~e,故也不能采用最短编码法。

最终选用字段编码法和直接控制法相结合的方法。将互斥的微命令安排在同一段内,兼容的微命令安排在不同的段内。b、i、j 这 3 个微命令是互斥的微命令,把它们安排在一个段内,e、f、h 这 3 个微命令也是互斥的,把它们也安排在另一个段内。此微指令的操作控制字段格式如图 6-31 所示。

其中:字段 1 的译码器输出对应的微命令为

00 无
01 b
10 i
11 j

字段 2 的译码器输出对应的微命令为

00 无
01 e
10 f
11 h

将全部 8 条微指令代码化可以得到

I_1 : 11100101
 I_2 : 10110010
 I_3 : 00000111
 I_4 : 01000000
 I_5 : 01011001
 I_6 : 10001111
 I_7 : 01100011
 I_8 : 10000111

21. 在微程序控制器中,微程序计数器 μPC 可以用具有加“1”功能的微地址寄存器 μMAR 来代替,试问程序计数器 PC 是否可以用具有加“1”功能的存储器地址寄存器 MAR 代替?

解: 在微程序控制器中不可以用 MAR 来代替 PC。因为控存中只有微指令,为了降低成本,可以用具有计数功能的微地址寄存器(μMAR)来代替 μPC 。而主存中既有指令又有数据,它们都以二进制代码形式出现,取指令和数据时地址的来源是不同的。

取指令: $(PC) \rightarrow MAR$

取数据: 地址形成部件 $\rightarrow MAR$

所以不能用 MAR 代替 PC。

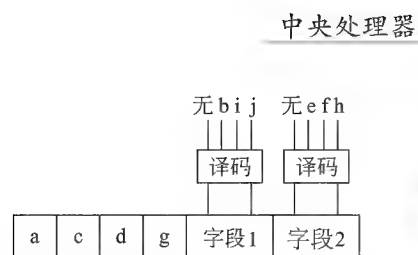


图 6 31 微指令控制字段格式

第 7 章

外部设备

7.1 基本内容要求

外部设备是计算机系统中不可缺少的重要组成部分,本章将介绍磁介质存储器的存储原理,常用磁介质存储设备和其他辅助存储设备,以及常见的输入输出设备的工作原理。

学习要求

- ◆ 了解外部设备的分类和外部设备的作用
- ◆ 了解磁介质存储器的记录介质和读写磁头
- ◆ 理解磁介质存储器的主要技术指标(记录密度,存储容量,平均存取时间,数据传输率)
- ◆ 掌握常见的数字磁记录方式(直接记录方法和按位编码方法),能画出数据序列的写电流波形
- ◆ 了解成组编码方式
- ◆ 理解硬盘上的信息分布形式
- ◆ 掌握磁盘地址的安排
- ◆ 掌握硬盘存储器技术参数的计算
- ◆ 了解硬盘的分区域记录技术
- ◆ 了解软磁盘存储器的特点
- ◆ 了解磁盘阵列(RAID)
- ◆ 了解光盘存储器的类型和工作原理
- ◆ 了解新型辅助存储器的特点
- ◆ 了解键盘的类型
- ◆ 理解非编码键盘的工作原理

- ◆ 了解其他输入设备的特点

- ◆ 了解印字输出设备的特点和分类
- ◆ 理解文本(字符)模式和图形模式的不同
- ◆ 理解点阵针式打印机的缓存和字库中存放信息的特点
- ◆ 了解显示器特点和分类
- ◆ 了解字符显示和图形显示的不同
- ◆ 理解字符显示器的显示缓存 VRAM 和字库中存放信息的特点
- ◆ 了解字符和图形显示器的同步控制

7.2 误点疑点解惑

1. 磁盘存储器的平均存取时间

在磁介质存储器中,当磁头接到读写命令,从原来的位置移动到指定位置,并开始读写操作的时间叫做存取时间。对于采用直接存取方式的磁盘存储器来说,平均存取时间包括3个部分:第一部分是指磁头从原先位置移动到目的磁道所需要的时间,称为平均寻道时间;第二部分是指寻道完成后等待被访问的信息旋转到磁头下方的时间,称为平均等待时间;第三部分是信息的读写操作时间,它与数据量、磁盘转速、记录密度、传输线的带宽等因素有关。通常读写操作时间取读扇区数据时间和传输数据时间两者中的最大值。严格说来,平均存取时间中还应当包括控制器开销(控制延时),即磁盘控制器从收到读磁盘命令到启动磁头移动的时间,控制延时一般很小。

例 7-1 设一个磁盘的平均寻道时间为 20ms,传输速率为 1MB/s,控制器延时是 2ms,转速为 5400r/min。求读写一个 512 字节的扇区的平均存取时间。

解: 磁盘转速为 5400r/min=90 转/s

平均等待时间为磁盘旋转半圈的时间,即

$$0.5 \text{ 转} \div 90 \text{ 转/s} = 0.0056\text{s} = 5.6\text{ms}$$

读写操作时间等于一个扇区的传输时间,即

$$0.5\text{KB} \div 1\text{MB/s} = 0.5\text{ms}$$

注意,这里假设了磁盘数据的读写速率高于数据的传输率。

所以,读写一个 512 字节的扇区的平均存取时间=平均寻道时间+平均等待时间+读写操作时间+控制器延时=20ms+5.6ms+0.5ms+2ms=28.1ms。

由于后两部分时间小于前两部分时间,所以经常可以将后两部分时间忽略不计,但要知道它们是实际存在着的。

2. 改进调频制提高记录密度的分析

调相制和调频制都属于位间无关型的按位编码。若将记录序列中相邻位联系起来,即采取位间相关性编码,可以进一步减少磁通翻转次数,从而提高记录密度。改进调频制

就是按照这个思路,从调频制改造得到的。

现将调频制的写电路波形重画于图 7-1 的上半部,首先分析哪些翻转需要保留,哪些翻转可以省略。写 1 时,位周期中间的翻转用来表示数据 1 的存在,因此它应当保留,但位周期边界处的翻转可以省略。连续两个 0 都没有位周期中间的翻转,所以它们的边界处应当有一个翻转,以产生同步信号。

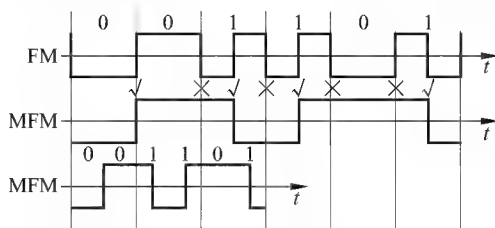


图 7-1 调频制和改进调频制的写电流波形分析

如图 7-1 所示,为记录相同代码,改进调频制的翻转次数约为调频制的一半。在相同技术条件下,改进调频制的位周期长度可以缩短为调频制的一半,使改进调频制的记录密度提高一倍。所以常将调频制称为单密度方式,将改进调频制称为双密度方式。

3. 硬盘存储器的圆柱面

硬盘中的信息分布涉及记录面、圆柱面、磁道、扇区等概念,其中圆柱面是一个需要特别关注的概念。硬盘往往是一个盘组,所有记录面上相同编号(位置)的诸磁道构成一个圆柱面。事实上,从物理意义上并没有圆柱面这一个实体,圆柱面数就等于一个记录面上的磁道数,圆柱面号就是对应的磁道号。

为什么要引入圆柱面的概念呢?磁盘上的信息通常是以文件的形式组织并且存储的,如果一个较长的文件在一条磁道存不完,是将它继续存放在同一记录面的相邻磁道上,还是将它继续存放在同一圆柱面的相邻记录面上?如果采用第一种方法,则更换磁道时必须进行寻道操作,这需要磁头沿半径方向运动,花费时间较长,且会有机械磨损;如果采用后一种方法,由于定位机构使所有记录面的磁头都对准同一序号磁道(处于同一圆柱面中),只需通过译码电路选取相邻盘面的磁头,即可继续读写,几乎没有时间延迟,也没有机械运动。很显然,应当采用第二种方法,让文件尽可能存储在同一圆柱面上,然后才是相邻圆柱面上。

4. 磁盘的基本操作

读写磁盘的基本操作可分成下面 3 个部分。

(1) 启动磁盘

主机用控制字启动磁盘。

(2) 寻址

根据主机发出的磁盘地址寻址,当已知某台号时寻找磁道、磁头和扇区。

① 回“0”道:当前磁头所在的磁道为当前道,也称现行道。复位时无论磁头在何磁道,都必须回到0道,即回到起始位置。

② 寻道:将当前道号与目的道号进行逻辑比较,两者符合表示已寻找到目的道号,两者不符合则将继续寻找。当目的道号大于当前道号,读写臂驱动磁头向内寻找(即向磁道号增大的方向移动);当目的道号小于当前磁道号,读写臂驱动磁头向外寻找(即向磁道号减小的方向移动)。寻道也称为定位,大约需要几十毫秒时间。

③ 寻(磁)头:这是通过译码电路实现的,因此速度较快,约需几十纳秒。

④ 寻找扇区:寻找扇区的时间取决于磁盘的旋转速度,速度越高,寻找时间越短。若寻找指定磁道的某一扇区,当磁头所在扇区恰好和地址码的扇区号一致,则扇区找到,寻找时间为0;当不一致时,磁盘最多需要旋转一圈才能找到。

(3) 磁盘的读写操作

根据读出操作控制字,在寻址完成后,即可将指定地址——某一扇区的信息从磁盘传送到主存中,称为读操作。

写操作是在写入操作控制字的控制下进行的。数据传送方向和读操作相反,它是将数据从主存传送到磁盘存储器中。

通常读写操作是在DMA控制器的控制下进行的,数据在主存和磁盘之间传送,以扇区为单位,不需要CPU干预。

5. 非编码键盘的行反转法

非编码键盘的按键一般排列成 m 行 \times n 列的矩阵形式,识别当前是否有键按下和判定当前按下的是哪一个键,均由软件来完成。对非编码键盘的识别方式通常有逐行扫描法、行反转法等。

行反转法又称作线反转法,其具体做法是,首先从行输出寄存器输出全“0”,然后从列输入寄存器输入8位数据,任何一个键按下时,输入的8位数据中肯定有一位为“0”,且按下的键肯定在这一列。接下来将行方向的输出寄存器改为输入寄存器,将列方向的输入寄存器改为输出寄存器,并且将刚才输入的8位数据(其中第 Y_i 位为“0”,其他位均为“1”)从列输出寄存器输出,再从行输入寄存器输入,则输入的8位数据中只有某一(X_i)位为“0”,其余位均为“1”。即可判定 X_i 和 Y_i 的交叉点上的键为当前按下的键。最后,经查表可得知当前的键值。

行反转法与逐行扫描法相比要简单一些。以一个 8×8 的键盘矩阵为例,采用逐行扫描法的扫描次数取决于按下键在矩阵中的位置,如果按下键位于第 X_0 行,则一次扫描就可以完成识别功能;如果按下键位于第 X_7 行,则需要扫描8次才能完成识别功能,显得有些繁琐。而行反转法在任何时候只要扫描一次,但是需要改变一次扫描方向,这是利用相关的并行接口芯片完全可以做到的。

6. 点阵针式打印机的打印方式

点阵针式打印机是一种串行击打式打印机,靠若干根钢针在字符点阵代码的控制下击打色带和纸,在纸面上印出与点阵代码相应的字符图案。

与显示器的显示方式类似,针式打印机也有两种打印方式。一种是文本字符方式,根据待打印字符的编码(存放在打印机缓存 RAM 中)从打印机字库 ROM 中依次取出字符的各列点阵数据,控制钢针在纸上打印出一个一个的字符。与字符显示方式不同的是,字库是按列组织字符点阵代码而不是按行组织的;并且点阵数据由 ROM 取出后,不经过并一串转换而直接送往打印头。另一种打印方式是点图形方式,将图形的点数据存入缓存 RAM 中,打印时从 RAM 取出点数据直接送打印头,驱动钢针打印出图形或汉字。

针式打印机虽然具有噪声大、印刷质量较差等缺点,但在打印大型宽行报表及需要多联打印的场合下,仍然具有其他非击打式打印机不可取代的优势。

需要说明的是,点阵式和针式是两个层面上的概念,点阵式打印不用字模产生字符,而是将字符以点阵形式存放在字库中。印字时,用取出的点阵代码控制在纸上打印出字符的点阵图形。点阵式打印组字灵活,可以打印各种字符、汉字、图形、表格等。针式打印机及所有非击打式打印机均采用点阵式打印,也就是说针式打印机肯定采用点阵式打印,但不能将点阵式打印与针式打印完全画等号。

7. 激光打印机的印字原理

激光打印机的核心技术就是所谓的电子成像技术,这种技术融合了影像学与电子学的原理和技术以生成图像,核心部分是一个可以感光的硒鼓。激光发射器所发射的激光照射在一个棱柱形反射镜上,随着反射镜的转动,光线从硒鼓的一端到另一端依次扫过。硒鼓是一只表面涂覆了有机材料的圆筒,预先带有电荷。计算机所发送来的数据信号控制着激光的发射,扫描在硒鼓表面的光线不断变化,有的地方受到照射,电阻变小,电荷消失;也有的地方没有光线射到,仍保留有电荷。最终,硒鼓表面就形成了由电荷组成的潜影。

碳粉是一种带电荷的细微塑料颗粒,其电荷与硒鼓表面的电荷极性相反,当带有电荷的硒鼓表面经过碳粉盒时,有电荷的部位就吸附了碳粉颗粒,潜影就变成了真正的影像。硒鼓转动的同时,另一组传动系统将打印纸送进来,经过一组电极,打印纸带上了与硒鼓表面极性相同但强得多的电荷,然后纸张经过带有碳粉的硒鼓,硒鼓表面的碳粉被吸引到打印纸上,图像就在纸张表面形成了。此时,碳粉和打印机仅仅是靠电荷的引力结合在一起,在打印纸被送出打印机之前,经过高温加热,塑料质的碳粉被熔化,在冷却过程中固定在纸张表面上。

将碳粉传给打印机之后,硒鼓表面继续旋转,经过一个清洁器,将剩余的碳粉去掉,以便进入下一个打印循环。

8. 图形和图像

图形和图像是现代显示技术中常用的术语,也是学生在学习过程中比较容易混淆的两个概念。图形(graphics)最初是指没有亮暗层次变化的线条图,如建筑、机械所用的工程图、电路图等。早期的图形显示和处理只是局限在二值化的范围,只能用线条的有无表示简单的图形。图像(image)则最初就是指具有亮暗层次的图,如自然景物、新闻照片等。经计算机处理后显示的图像称作数字图像,就是将图片上连续的亮暗变化变换为离散的数字量,且以点阵列的形式显示输出。

在显示屏幕上,图形和图像都是由称作像素的光点组成的。光点的多少称作分辨率,光点的深浅变化称作灰度级(在单色显示器上表现为灰度级,在彩色显示器上表现为颜色)。分辨率和灰度级决定所显示图的质量。高分辨率和多灰度级的光栅扫描的显示器不仅可以显示图像,也可以显示图形。现在的图形也可以有颜色、深浅层次的变化。但是,图形学和数字图像处理是两个不同的学科,它们研究的问题是不同,应用领域不同,使用的技术方法不同,图形和图像的输入手段也不同。

图形学的主要任务是研究如何用计算机表示现实世界的各种事物,并且形象逼真地加以显示,如动画设计、花布图案设计、地图的显示等平面图,飞机、汽车、建筑物的造型设计等立体图,这些图的显示效果要有真实感,需要有深浅和颜色。图形学所用的技术包括点、线、面、体等平面和立体图的表示和生成。由于要在平面上显示立体图,还要研究阴影的产生,隐藏线、隐藏面的消除技术以及光照方向与颜色的模拟等技术。

数字图像处理所处理的对象多半来自客观世界,例如,由摄像机摄取下来存入计算机的数字图像(遥感图像、医用图像等)。图像和图形相比,由于后者可以按人的意志描绘,所以无噪音干扰,而且规则整齐,富有创造性;前者则可能充满噪音,图像很不清晰。由于摄取的位置随机,图像可能发生畸变。图像处理的任务是去除噪音,恢复原形,使图像清晰,并且从中抽取有用的信息,以供观察。

图像主要用摄像机输入,经数字化以后逐点存储,因此,图像需要占用非常庞大的主存空间。而在计算机中表示图形,则只需存储绘图命令和坐标点,没有必要存储每个像素点。

9. CRT 显示器的有关技术指标

在选择 CRT 显示器时,通常会涉及到它的一些主要技术指标,下面讨论几个容易引发错误的技术指标。

(1) 点距

点距是指 CRT 上同一像素中两个颜色相同的磷光粉像素之间的距离。点距越小,显示器画面就越清晰细腻、自然,分辨率和图像质量也就越高。图 7-2 所示的便是点距、水平点距和垂直点距之间的关系。其中,0.28 是点距,0.24 是水

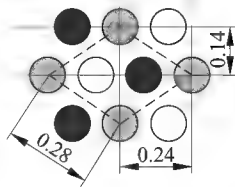


图 7 2 点距、水平点距和垂直点距之间的关系

平点距,0.14 是垂直点距。

很明显,点距要大于水平点距和垂直点距。

(2) 视频带宽

这是表示显示器显示能力的一个综合指标,它能够决定显示器性能的好坏以及一台显示器可以处理的信息量。视频带宽指每秒钟电子枪扫过的图像点的个数,即单位时间内每条扫描线上显示的点数的总和。对于 17 英寸的显示器而言,1600×1200 的最高分辨率可能没有什么实际意义,但是在 1024×1024 分辨率下,带宽 110MHz 的显示器和带宽 200MHz 的显示器的差异是很明显的。

带宽的大小是有一定计算方法的,用户在选择一款显示器的时候,可以根据一些参数来计算显示器的带宽,或者根据带宽来计算一些参数。其计算公式为:

$$\text{视频带宽} = \text{行数} \times \text{列数} \times \text{刷新率} \times 1.3$$

例如,一台显示器在 1280×768 分辨率和 85Hz 刷新频率下正常显示时,可以计算出显示器的视频带宽 $=1280 \times 768 \times 85 \times 1.3 = 87\text{MHz}$ 。当然,也可以根据显示器的带宽计算出显示器在最大分辨率下的刷新频率等参数。与行频相比,视频带宽更具有综合性,也更能直接反映显示器的性能。

视频带宽越大表明显示器显示控制能力越强,显示效果越佳。在同样分辨率下,视频带宽高的显示器不仅可以提供更高的刷新频率,而且在画面细节的表现方面往往更加准确清晰。视频带宽决定着显示器的分辨率和刷新频率,应该说是带宽越大越好。

低档显示器的视频带宽多为 110MHz,甚至更低;中档产品的可以达到 135~160MHz;而高档产品则可以达到 200MHz,甚至更高。

10. VRAM 的容量和内容

显示缓存区又称视频随机存储器 VRAM,显示器一方面对屏幕进行光栅扫描,一方面同步地从 VRAM 中读取显示内容,送往显示器件。因此,对 VRAM 的操作是显示器工作的软、硬件界面所在。为了在指定的屏幕位置显示某个字符,就需向 VRAM 的相应单元写入该字符编码;为了更新屏幕显示的内容,就需相应地刷新 VRAM 的内容;为了使画面呈现某种动画效果,就需要使 VRAM 中的内容作相应的变化,或者在读取时进行某种地址转换。

VRAM 一般设置在显示器控制器(显卡)上。在微型计算机中,VRAM 占主存空间,从软件上讲它可以视为主存的一部分;在独立的显示终端中,VRAM 作为外设存在,与主存分离。

当用字符方式显示时,VRAM 中的内容一般包含显示内容和显示属性两个部分。前者提供显示字符代码,后者提供有关显示的属性信息。这两部分可以分别存放在两个缓冲存储器中,一个称为基本显示缓存,另一个称为显示属性缓存。通常将这两个存储体统一编址,一个为偶数地址,另一个为奇数地址;也可以将两部分存放在一个缓存中,依靠地

址码为偶数或奇数进行区分。基本显示缓存中存放的是一帧待显示字符的 ASCII 码或其他形式的编码,字符的点阵信息则放在字库 ROM 中。在这种方式下,一个字符编码占缓存的一个字节,因此缓存的最小容量是由屏幕上字符显示的行列规格决定的。显示属性缓存的容量应当与基本显示缓存的容量相同。

如果采用图形方式显示,VRAM 中的内容就是一帧待显示的图形的像点信息,其代码“1”和“0”分别表示图形中的亮点和暗点。这些图形可以是几何图形、任意曲线图形、汉字或字符。这里需要特别说明的是,在图形方式下字符的点阵是以位图的形式直接存放在显示缓存中的,因此字符可以像素为单位在屏幕的任意位置上显示。在图形方式中,缓存的容量不仅取决于屏幕分辨率的高低,还与显示的颜色种类有关。在单色显示时,图形的每个点一般只用一位二进制代码来表示,在彩色显示时,每个点需要由若干位代码来表示。

11. 字符显示原理和具体显示过程

字符显示常采用光栅扫描法,它以点阵为基础,将欲显示字符分解成 $m \times n$ 个点组成的矩阵,并将能显示的所有字符的点阵存入由 ROM 构成的字符发生器(字库)中。字符点阵的多少取决于对显示字符的质量要求和字符块的大小。字符块是指在显示屏幕上每个字符所占的点数,通常称为“字符窗口”,它应包含字符本身所占点阵和字符之间的间隔所占点阵,显然,每个字符窗口所占点阵数越多,显示的字符越清晰,显示质量越高。

一般的字符显示屏幕上可显示 80 列 \times 25 行共 2000 个字符,即有 2000 个字符窗口。在单色字符显示器中,常用的字符窗口为 9×14 点阵,字符本身只占 7×9 点阵,字符“A”在字符窗口中的位置如图 7-3(a)所示。从图中可以看出,每个字符窗口包含 14 个点阵字节,对于任何字符来说,各自的点阵字节是固定不变的,它们事先被存放在只读的字符发生器中,每个字符在字符发生器中占用 14 个字节,例如,“A”字符存放在字符发生器中的 14 个点阵字节(行点阵码)为:10H、28H、44H、82H、82H、82H、FEH、82H、82H、00H、00H、00H、00H、00H,每个点阵字节对应的地址为 12 位,高 8 位为“A”字符的 ASCII 码,低 4 位为字符点阵的行号(0000B~1101B)。对于字符“A”来说,它的点阵字节应存放在字符发生器中从“410H”地址开始的 14 个连续地址中,如图 7-3(b)所示。

由于每个字符或符号的点阵字节是不同的,但又是固定不变的,所以字符发生器一般用 ROM 构成,其容量必须能存放可在屏幕上显示的所有的字符或符号的点阵字节,而且每个字符或符号在字符发生器中的地址码由 12 位二进制数构成。因此,只要知道当前要显示的是什么字符,便可以从字符发生器中找到该字符的点阵字节。

在显示屏幕上每个字符行一般要显示多个字符,最多可以达到 80 个字符,为了在扫描过程中能及时获得各个字符窗口需显示的字符,应将这些欲显示字符的 ASCII 码预先存入 VRAM 中。字符显示器的 VRAM 分成两部分,一部分用来存放显示字符的 ASCII 码,每个字符占一个字节,另一部分用来存放显示属性。在单色显示器中,显示属性一般

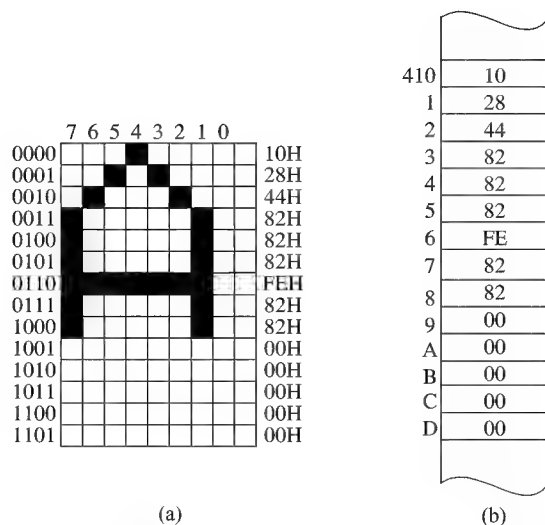


图 7-3 字符“A”的点阵位置和行点阵的存放

包括显示色、底色,是否增辉(加亮)、是否闪烁等。彩色显示器中,显示属性还应表明颜色的类型等。

采用光栅扫描方式显示字符时,并不是对每个字符单独扫描,而是对一行上的所有字符的同一条扫描线上的点阵进行扫描,对于 9×14 的字符窗口,只有扫完了 14 条扫描线,这一行上的所有字符才会完整地显示在显示屏上。

例如,要求在屏幕的第 0 行的第 0~5 个字符窗口显示“HELLO !”这 6 个字符和符号,而其他字符窗口均为空白区。于是 VRAM 中的内容应为它们的 ASCII 码,如图 7-4 所示。

显示的具体过程如下:从字符发生器的 0480H、0450H、04C0H、04C0H、04F0H、0210H、0200H...0200H 共 80 个地址中的第一个点阵字节去控制电子束完成第一条扫描线的扫描,然后继续从字符发生器的 0481H、0451H、04C1H、04C1H、04F1H、0211H、0201H...0201H 共 80 个地址中的第二个点阵字节去控制电子束进行第二条扫描线的扫描,上述操作重复 14 次,即可完成 14 行的扫描,于是“HELLO !”这 6 个字符便清晰地显示在屏幕上了。

12. 显示器的同步控制

不论字符显示还是图形显示,都要求行、场扫描和视频信号的发送在时间上完全同步,即当电子束扫描到某字符或某像素的位

VRAM	
0	48
1	45
2	4C
3	4C
4	4F
5	21
6	20
⋮	
1999	20

图 7-4 VRAM 中的内容

置时,相应的视频信号必须同时输出。为此,在 CRT 显示器中设置几个计数器,对显示器的主频脉冲进行分频,产生各种时序信号,控制对 VRAM 的访问,对 CRT 的水平扫描和垂直扫描,以及视频信号的产生等。

字符方式和图形方式下对计数器的设置是有区别的,下面分别加以讨论。

(1) 字符显示的同步控制

以单色字符显示器为例,每帧最多显示 25 行 \times 80 列字符,字符窗口为 9×14 ,其中字符本身占 7×9 点阵。字符显示器的定时控制电路中设置了点计数器、字计数器(水平地址计数器)、行计数器和排计数器(垂直地址计数器),由它们来控制显示器的逐点、逐字、逐行、逐屏幕的刷新显示。为了避免扫描行和字符行这两个概念的混淆,在此把扫描行仍称为行,而把字符行称为排。

① 点计数器分频 $9:1$

设置点计数器的目的是为了提供下列控制信号:读 VRAM,控制一个字符区间内的横向间隔消隐,对字计数器计数。

每个字符点阵横向 7 个点,间隔 2 个点。点脉冲一方面控制视频信号产生像点,一方面对点计数器计数。每计数 9 个点,完成一次计数循环,分频关系 $9:1$ 。

每次访问 VRAM,读出一个显示字符的编码。以字符编码为高位地址,以扫描行号为低位地址,访问字符发生器 ROM,从中读出 7 位代码。由点脉冲控制时间,在屏幕的一行扫描线上一次显示 7 个像点(亮或暗)及 2 点间隔。每当点计数器完成一次计数循环后,就应访问一次 VRAM,以读取下一次显示字符的编码,同时向下一级的字计数器提供一个计数脉冲。

② 字计数器分频 $(80+L):1$

设置字计数器的目的是为了提供下列控制信号:VRAM 低位地址信息,控制一条水平扫描线内的显示与消隐,向显示头提供水平同步信号,对行计数器计数。

字计数器计数一次,导致一次正程扫描,而一行水平扫描线包含 80 个字符的显示区间。回扫与线性度不好的边缘部分应当消隐,将它们折合成 L 个字符位置, L 的值与显示头制造技术有关。因此,字计数器计数 $(80+L)$ 之后,完成一次计数循环,分频关系 $(80+L):1$ 。每完成一次计数循环,产生一次水平同步信号,启动下一次水平扫描,且使下一级的行计数器计数一次。

访问 VRAM 的地址,取决于该字符在屏幕上的显示位置(行、列号)。字计数器提供的当前显示位置列号,可以作为产生 VRAM 低位地址的依据。

③ 行计数器分频 $(9+5):1$

设置行计数器的目的是为了提供下列控制信号:访问字符发生器 ROM 的低位地址,控制一排字符中哪些扫描行显示,哪些扫描行消隐,控制光标显示,对排计数器计数。

每完成一次水平扫描,行计数器计数一次。一排字符占 9 行水平扫描线,然后是作为

排间间隔的 5 行水平扫描线。所以行计数器计数 14 次之后,完成一个计数循环,分频关系 $14:1$ 。每完成一次计数循环,对下一级的排计数器计数,启动新的一排字符显示。

④ 排计数器分频 $(25+M):1$

设置排计数器的目的是为了提供下列控制信号:VRAM 高位地址信息,向显示头提供垂直同步信号,控制一场显示过程中的显示段与消隐段。

每显示完一排字符,行计数器完成一次计数循环,则排计数器计数一次,导致一次自上而下的正程扫描,可以显示 25 排字符。回扫和线性度不好的边缘部分应当消隐,折合为 M 排, M 的值与显示头制造技术有关。因此,排计数器计数 $(25+M)$ 次之后,完成一个计数循环,分频关系 $(25+M):1$,相应地实现一场的显示,发出一次垂直同步信号。

读 VRAM 时,排计数器值决定字符的行号,可以作为高位地址的依据。

综上所述,将字符显示的同步控制关系做一简单小结:

- 点计数器一个循环,访问 VRAM 一次以读取显示字符的编码,VRAM 的地址根据排计数器和字计数器值决定。
- 每读一次 VRAM,就紧跟着读一次字符发生器 ROM,由 VRAM 读出的字符编码产生 ROM 的高位地址,行计数器值决定 ROM 的低位地址。
- 每次从 ROM 读出显示字符的一行 7 位行点阵码,由点脉冲控制逐位显示 7 点。
- 由于每条扫描线只能显示一排字符(80 列)的一行,所以上述访问过程需要重复 9 遍(每遍又要多次访问 VRAM 与 ROM,以读取不同字符),才能显示完整的一排字符。
- 字计数器循环一次,发一次水平同步信号。
- 排计数器循环一次,发一次垂直同步信号。

(2) 图形显示的同步控制

以分辨率为 640×480 的图形显示器为例,图形显示器的定时控制电路中设置了点计数器、列计数器和行计数器。

① 点计数器分频 $8:1$

图形以像素为单位,在 VRAM 中以字节为单位按地址存储,即将一条水平扫描线自左向右,每 8 个点的代码作为一个字节,存放在一个编址单元中。因此点脉冲经点计数器 8 分频之后产生一个脉冲,使列计数器计数,并访问一次 VRAM,读出一个字节(8 个点)。

② 列计数器分频 $(80+L):1$

列计数器又称字节计数器。光栅从左向右扫描一行,正程显示 80 字节共 640 点。列计数器所附加的 L 次计数,作为行线逆程回扫时间,逆程回扫应当消隐。

③ 行计数器分频 $(480+M):1$

行计数器的一次计数循环实现一场显示,其中 480 次计数,对应于场正程扫描,显示 480 行,附加 M 次计数,对应于场逆程回扫,逆程回扫应当消隐。

行计数值与列计数值决定屏幕当前显示位置(8点一组),相应的 VRAM 地址为:行号 $\times 80$ +列号。列计数器计数一个循环,输出一个行扫描(水平)同步信号;行计数器计数一个循环,输出一个场扫描(垂直)同步信号。

7.3 相关知识介绍

1. 磁记录方式的性能特点

为了比较各种记录方式的性能,下面通过参数进行分析。

(1) 自同步能力

自同步能力是指能否从单个磁道读出的脉冲序列中提取同步信号的能力。能直接提取同步信号称为有自同步能力,否则称为无自同步能力。显然,只有读出的序列是呈周期性的,才可能从规定的位周期中提取出同步信号。自同步能力的强弱可以用最小磁通翻转间隔和最大磁通翻转间隔的比值 R 来衡量。 R 值越大,自同步能力越强。对于无自同步能力的记录方式,必须设立专门的时钟磁道,称为外同步。例如, NRZ 制和 NRZ-1 制无自同步能力, RZ、PM、FM、MFM、 M^2 FM 制具有自同步能力,其中 FM、MFM 制的 $R=0.5$, M^2 FM 制的 $R=0.4$ 。

(2) 编码效率

编码效率是指位密度与最大磁通翻转密度之比,或者是每次磁层翻转所能记录数据位数的多少。NRZ、NRZ-1、MFM、 M^2 FM 制记录 1 位二进制信息最多翻转一次,故编码效率为 100%,而 RZ、PE、FM 制记录 1 位二进制信息最大翻转次数为 2,故编码效率为 50%。

显然,编码效率越高,记录密度就越高。从这个意义上, NRZ 制和 NRZ-1 制可获得高的记录密度,但因为它们不具备自同步能力,需要设置专用的同步磁道来产生外同步信号,所以并不能实现高密度的记录。编码效率的提高,不仅可提高记录密度,而且可减少噪音抖动,增加抗干扰能力。

(3) 读出分辨率

读出分辨率是指磁记录设备对读出信号的分辨能力,也就是指每次磁化翻转可判别信息的能力。

通常在读出过程中采用峰值鉴别法,设置一个检读窗口。如果在窗口范围内检测到峰值,则这一位是“1”。当某一位峰值偏离到窗口以外时,这个窗口无法检测到,而被邻位窗口所检测,将产生读出误差。若检读窗口大,则允许读出脉冲有较大的抖动。由此可见,检读窗口宽度大的记录方式,对读出信号峰值超前或滞后都能检读出来,说明其具有较高的读出分辨率。

对于 NRZ 制和 NRZ-1 制记录方式,每一位数据仅需检读一次,即检读窗口宽度等于

T_0 , 而 PM、FM、MFM、 M^2 FM 制每一位数据需检读两次, 因为磁通翻转可能发生在位周期中间也可能发生在边界上, 检读窗口宽度为 $0.5 T_0$ 。

(4) 信息的独立性

读出时, 如果某一位信息出现误码, 不会影响到后续其他信息位的正确性, 这叫做信息的独立性。反之, 称为误码传播。

NRZ 制容易造成误码传播, 若漏读 1 位, 则以后各位将会出错, “1”误为“0”, “0”误为“1”。见图 7-5。

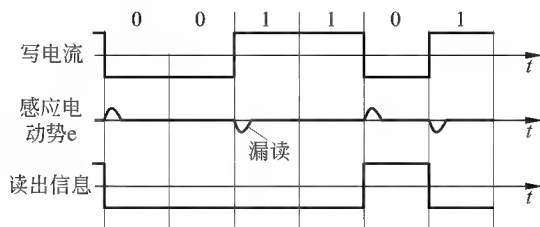


图 7 5 NRZ 制的误码传播

PE 制也会造成误码传播, 若漏读一个“1”, 会波及以后各位, 使一连串“1”被误作为“0”, 直至真正的“0”出现为止, 如图 7-6 所示。图中 d 代表数据位的读出波形, c 代表位周期起始处的读出波形。假定图 7-6 中所示的第二位数据漏读, 则第三位的 c 将被作为 d 读出, 下一次收到的 c 将作为 d 读出, 被误作为“0”读出, 由此波及以后位, 直至出现真正的“0”为止。

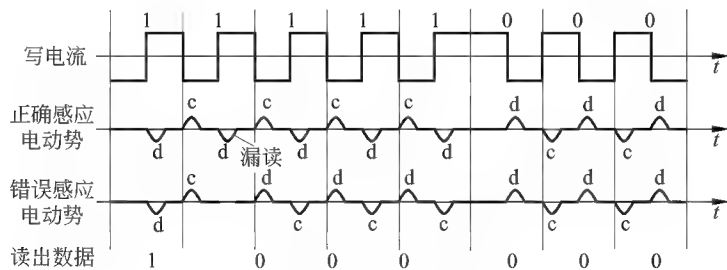


图 7 6 PE 制的误码传播

除去上面提到的性能参数以外, 还有像信道带宽、抗干扰能力、编码译码电路的复杂性等因素, 都对记录方式的取舍评价产生影响, 在这里就不一一讨论了。总之, 所选择的记录方式应尽量做到:

- ① 具有较强的自同步能力;
- ② 有较高的编码效率, 以提高记录密度;
- ③ 有较宽的检读窗口, 以提高读出分辨能力;

- ④ 有较强的抗干扰能力,以避免误码传播;
- ⑤ 编码、译码电路成本低,容易实现。

2. 几种磁记录方式的读出过程分析

磁介质存储器在读出时,每当磁通的翻转位置经过读磁头的下方时将产生感应电动势,下面分析几种常见的磁记录方式的读出过程。

(1) 不归零-1制(NRZ-1)

读出时,逢“0”没有读出信号,逢“1”就有读出的感应电动势。由于 NRZ-1 制没有自同步能力,需要外加同步信号来识别各个位周期,所以 NRZ-1 制不能直接用于像磁盘这种单道记录方式中,但可用于像磁带这种同时读写多道的设备之中。有两种方法来产生外同步信号:一种是在磁带上专门写入一个同步信号道,每位均为“1”,即每位均有一次磁通翻转,读出时每位都产生一个同步信号,用以选通数据道的各位;另一种方法是不设专门的同步道,而是让同时读出的各位(称为一个带字)采取奇校验,则每个带字中至少有一个“1”,可以提取出来作为同步信号。

采取外同步的方法限制了记录密度的提高,这是因为磁带在运动中难免存在扭斜,各位并不总是准确地同在一根垂直线上(与磁带运动方向垂直)。当记录密度较高时,外同步信号就难以准确地选通其他各位。NRZ-1 曾直接应用在早期的低速磁带机中,现在它仍是多种记录方式的基础或中间形式。

(2) 调相制(PE)

读出时,位周期的中央产生的感应电动势既是数据信号,也是同步信号;位周期交界处可能产生的感应电动势被弃之不用。根据读出信号的相位(感应电动势的正负),可以识别出该位信息是“0”还是“1”。

(3) 调频制(FM)

读出时,每个磁化翻转区都将产生一个感应电动势,所以读出信号序列中包含同步信号和数据信号。通过分离电路,将每个位单元起始处的信号分离出来,作为该位的同步信号。它将触发一个单稳电路,宽度为 $\frac{2}{3}T$ (T 为位周期宽度),形成一个选通窗口,用以选通位周期中部的读出信号,这个读出信号就是数据信号。

3. 群码制(GCR)

群码制是一种成组编码方式,常用的 GCR(4,5)是一种广泛用于高密度数字磁带机上的记录方式,它的基本方法是 将 4 位一组的数据码,整体转换成 5 位一组的记录码;在数据码中连续 0 的个数不受限制,但在转换后的记录码中,连续 0 的个数不超过两个;将转换后的记录码按 NRZ-1 制记入磁带。从信息量的角度看,从 4 位扩大为 5 位,组合数增加了一倍,可以只选取其中连续 0 的个数不超过 2 的组合,将连续 0 的个数在 2 以上的组合丢弃不用。变换规则如按表 7-1 所示。

表 7-1 GCR(4,5)变换规则

数据码	记录码	数据码	记录码
0000	11001	1000	11010
0001	11011	1001	01001
0010	10010	1010	01010
0011	10011	1011	01011
0100	11101	1100	11110
0101	10101	1101	01101
0110	10110	1110	01110
0111	10111	1111	01111

GCR(4,5)属于游程长度受限码。数据码长度 $m=4$, 记录码长度 $n=5$, 在记录序列中两个“1”之间至少存在“0”的个数 $d=0$, 最多存在“0”的个数 $k=2$, 一次变换的最大数据长度与最小数据长度之比值 $r=1$ 。

4. 1/4 英寸的数据流磁带机

磁带存储器有许多种, 其中价格最便宜的当属 1/4 英寸的数据流盒式磁带机 QIC (quarter inch cartridge)。数据流磁带机的容量通常为 1GB 以上, 当采用数据压缩技术后, 磁带机的容量可以增加 1 倍。

数据流磁带机是将数据连续地写在磁带上, 每个数据块间插入记录间隙, 使磁带机在数据块之间不启停, 从而简化了磁带机的结构。

传统的启停式磁带机采用多位并行读写, 读写磁头随磁道增加而增加。而数据流磁带机的读写磁头只有一个或两个, 采用类似于磁盘的串行读写方式。

以 4 道数据流磁带机为例, 4 个磁道的排列次序如图 7-7 所示。当记录信息时, 先从第 0 道的带头 BOT 开始, 记到带尾 EOT; 然后又从第 1 道的带尾 EOT 反向记录到带头 BOT; 第 2 道又从 BOT 到 EOT, 第 3 道则从 EOT 到 BOT。读出时也按这个顺序, 这种方式称为蛇形记录方式。与一般磁带机相比, 这种蛇形记录方式节约了数据读写过程中的倒带时间, 使后援时间大为缩短。

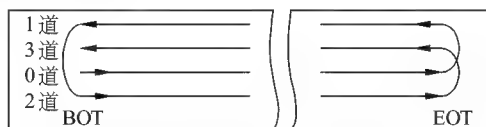


图 7-7 4 道 1/4 英寸磁带蛇形串行记录方式

磁带存储器的发展趋势主要是提高记录密度, 提高数据传输率和提高可靠性。

5. CD-ROM 读盘方式

CD-ROM 读盘方式有以下几种。

恒定线速度(constant linear velocity, CLV):这是早期光驱使用的读盘方式,多用于8倍速以下的光驱。这种光驱在运行时,总是以一定的线速度来运转,这样在读取内圈数据的时候,由于半径小,光驱就要加大主轴电机的马力来提高转速以获得与外圈相同的线速度。即读内圈数据时光驱转速高,读外圈数据时光驱转速低。随着光驱速度的不断提高,电机频繁地改变转速势必会大幅度缩短寿命,因此恒定线速度无法适应高倍速光驱。

恒定角速度(constant angular velocity, CAV):光驱在运行的时候,不论是读取外圈数据还是读取内圈数据,主轴电机的转速都恒定不变。这样电机就不用来回的改变转速,从而提高了光驱的寿命。采用这种技术的光驱在读取外圈数据和内圈数据时的传输率不同,读取内圈数据时传输率较低,读取外圈数据时传输率较高。CAV的优点是读取速度快,但对一些对速度要求不高的盘片的读取精度和纠错度不如CLV,目前,这种方式被大多数的光驱所采用。

局部恒定角速度(partial-constant angular velocity, P-CAV):这种方式将CAV和CLV合二为一,先保持盘片转动的角速度不变,因此读取速度会随着激光头往盘片外圈的移动而逐渐加快,到达某一速度后(稳定工作的极限速度),则切换到以线速度恒定的方式读取,此时读取速度固定而转速则会慢慢下降。P-CAV方式通过智能软件识别盘片,自动在CAV和CLV技术之间切换,兼顾速度和读取精度,达到最理想的读盘速度。

6. 光盘刻录机工作原理

在光存储盘片的表面有一层薄膜,大功率的激光照射在这层薄膜上时,薄膜上会形成平面和凹坑,光盘读取设备将这些平面和凹坑信息转化为“0”和“1”,将光盘上的物理信息转换为数字信息。对于CD-R盘片,这种薄膜上的物理变化是一次性的,因此CD-R盘片只能写入一次,不能重复写入。而CD-RW盘片上的薄膜材质多为银、硒或碲的结晶体,这种薄膜能够呈现出结晶和非结晶两种状态,在激光束的照射下,可以在两种状态之间转换,所以CD-RW盘片可以重复写入。

(1) 刻录机的缓存

为保证刻录质量,高速刻录时除了对盘片的要求比较高以外,缓存大小也十分重要。从理论上讲,缓存越大,刻录失败率则越低。但限于成本,一般刻录机缓存为2MB。在刻录开始前,刻录机需要先将一部分数据载入到缓存中,刻录过程中不断从缓存中读取数据刻录到盘片上,同时缓存中的数据也在不断补充。一旦数据传送到缓存里的速度低于刻录机的刻录速度,缓存中的数据就会减少,缓存完全清空之后,就会发生缓存欠载问题,导致盘片报废。所以在没有防刻死技术的刻录机上缓存大小直接影响到刻录的成功。缓存越大,则发生缓存欠载问题的可能性就越低。

现在,一般的刻录机中都安置了缓存。缓存就像一个水库,将上游即数据源来的数据暂时囤积起来,之后以一定的速度供给刻录系统,并在上游暂时断流时将囤积的数据继续向刻录系统供给,避免刻录系统也随之断流。缓存容量越大,则发生刻死现象的可能性就

越小。

(2) 刻录模式

刻录机的刻录模式与光驱的读盘模式差不多,刻录模式主要有以下4种:恒定线速度(CLV)、恒定角速度(CAV)、局部恒定角速度(P-CAV)、区域恒定线速度(Z-CLV)。

CLV模式的刻录速度稳定,激光可以固定的功率来刻录盘片,能够保证刻录品质,但不适合高速的刻录机。

CAV模式的转速不变,读速(传输率)逐渐提高,改变传输率就意味着改变激光功率,所以现在基本上不采用该模式进行刻录。

P-CAV模式是CAV和CLV的合二为一,可以理解为从转速不变读速逐渐提高到读速不变转速逐渐减小。

而Z-CLV(Zone-CLV)模式是目前市场上大多数高速刻录机采用的刻录模式,其原理是将盘片由内圈到外圈分成数个区域,在每一个区域用稳定的CLV速度进行刻录,在区段与区段之间采用CAV模式过渡,逐步提升速度。即在第一段起读速不变,转速逐渐减小;第二段起读速不变,转速逐渐减小,……,直至到刻录机的标称速度为止。这样做的好处是减短了刻录时间,并能确保刻录品质,只是在此模式下,每一次切换速度时刻录过程都会有明显的中断。

(3) 刻录保护技术

刻录机在没有考虑保护技术之前,刻录时经常会出现缓存欠载现象。随着刻录机写入速度的不断提升,单靠旧有技术简单的增加内建缓存已经不能有效地解决刻录过程中的刻死问题,而且缓存也不可能无限地增加下去,于是众多新的刻录保护技术应运而生。

比较知名的刻录保护技术有两种:第一种技术是刻录机实时监测着硬件缓存,当因为某种数据供给原因使缓存中数据量减少,低于警戒水平时,暂时中断刻录进程,使系统等待缓存中积累起足够多的数据后再继续进行刻录工作,而继续刻录已不可能完全连接上以前的部分,这就出现了“Link”区(两次续刻间隔区)的长度问题。第二种技术是检测缓存中有效数据量,当其低于标准时,暂停刻录过程并储存终点,在此状态等待缓冲区的充盈后,检测上一次结束终点,进行新的刻录过程。

7. 无线鼠标

长长的鼠标线的确是有些麻烦,它严重的束缚着用户的使用。为了摆脱这种烦恼和干扰,无线鼠标自然就诞生了。无线鼠标又分为红外线鼠标和射频鼠标两种。

(1) 红外线鼠标

红外线的特点是没有穿透能力,也就是说红外线的发射器和接收器之间不能有任何障碍物,这就大大降低了无线连接的方便性。此外,红外线的特性也决定了其接收信号的距离很小,一般在1m以内,而且角度范围也很小。但是由于红外线技术相当普及,且生产成本低,因此目前市场上仍有不少百元左右的红外线鼠标。

(2) 射频鼠标

相对红外线鼠标而言,射频鼠标则要先进得多,也是目前的主流。它具有 360°的信号接收角度与 3m 左右的接收距离,而且鼠标与射频发射器之间不受任何普通物体干扰。此外,射频鼠标的抗干扰能力很强,绝对不会与周围的无线鼠标冲突,这些都是红外线鼠标所不具备的。除此之外,射频鼠标的功耗也要比红外线鼠标小不少。

8. 鼓式宽行打印机

行式打印机有窄行和宽行之分。宽行打印机每行可包含 80、120、132 或 160 个字符。其特点是打印速度快,一般每分钟可打印 600~1200 行,最快的可达到 2000 行。宽行打印机有鼓式、带式 and 链式之分,下面仅讨论鼓式宽行打印机,其结构原理图如图 7-8 所示。

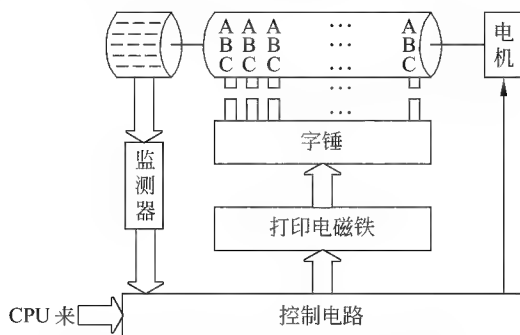


图 7-8 鼓式宽行打印机结构原理图

鼓式宽行打印机把可打印的字符铸在一个旋转的字鼓上,字鼓是一个圆柱体,在其表面上沿圆周方向均匀地刻着多种字符,沿轴线方向同一行上是相同的字符。如果鼓式宽行打印机的行宽为 80,则字鼓表面每行就刻有 80 个字符。

字鼓被一个金属罩所包围,而仅仅在打印位置上留有一个缝隙。在打印位置下面设置有一排字锤(80 个),每个字锤与字鼓上的一列字符对应,并由相应的打印电磁铁驱动,打印锤平时不与字鼓接触。打印机工作时,字鼓由电机带动匀速旋转,当被选字符处于字锤之下时,相应字锤被电磁铁驱动击打,实现打印一个字符功能,字轮每旋转一圈,打印完一行字符。

为了能在任何时间都知道字鼓上哪种字符正处于打印位置,通常与字鼓同轴安装有一个编码盘。编码盘与字鼓同步旋转,编码盘上的编码与字鼓上同一行字符的 ASCII 码相对应。例如,若字鼓上这一行是字符“A”,那么编码盘上同一行编码为 41H。需要打印时,首先将一行欲打印字符通过接口置入打印机的控制电路中,接着将这一行中各字符与编码盘上的第一组编码比较,比较相同的电磁铁通电,被吸动的字锤击打,即可实现多个字锤并行打印相同的字符。随着字鼓的旋转,重复上述过程直到字鼓转完一圈,一行字符被打印完毕。

例如,要求打印的一行字符为“COMPUTER ORGANIZATION”,其打印过程如下:

次数	打印纸上				备注
1			A	A	打印字符 A
3	C		A	A	打印字符 C
5	C	E	A	A	打印字符 E
7	C	E	GA	A	打印字符 G
9	C	E	GA I	A I	打印字符 I
13	C M	E	GA I	A I	打印字符 M
14	C M	E	GANI	A I N	打印字符 N
15	COM	E	O GANI	A ION	打印字符 O
16	COMP	E	O GANI	A ION	打印字符 P
18	COMP	ER	ORGANI	A ION	打印字符 R
20	COMP	TER	ORGANI	ATION	打印字符 T
21	COMPUTER		ORGANI	ATION	打印字符 U
26	COMPUTER	ORGANIZ		ATION	打印字符 Z

当字鼓旋转到字符“Z”之后,完成了本例要打印的一行字符。但只有当字鼓旋转一圈,所有的字符都通过了字锤位置后,才能打印下一行字符。

鼓式宽行打印机虽然打印速度快,但打印机的机械结构比较复杂,要求精度高,可打印的字符越多,字鼓就越大,而且一旦制造完毕,字符不可更改、增加或删除,无法打印图形和汉字。

9. 色光三原色和颜料三原色

自然界中有着各种各样的颜色,都是通过光反映给人们的。而这些色彩几乎都可以由选定的 3 种单色光以适当的比例混合得到,而且绝大多数的彩色光也可以分解成特定的 3 种单色光。这 3 种选定的颜色被称为三原色。

三原色分为两类,一类是色光三原色,又称加色法三原色,是将 3 种加性原色——红、绿、蓝线性叠加后组合而成;另一类为颜料三原色,又称为减色法三原色,将 3 种减性原色——青(所有红光被吸收)、黄(所有蓝光被吸收)、品红(所有绿光被吸收)线性叠加后组合而成。

色光三原色是光色混合。光色的混合为加色混合,是光线的增加,两种色光混合,光度为两色之和,合色愈多,则光度愈强,愈近于白。通过科学实验表明,人们的眼睛对红、绿、蓝(分别用字母 R、G、B 表示)3 种颜色反应最灵敏,而且它们的配色范围比较广,用这 3 种颜色可以任意配出自然界中的大部分颜色,彩色显示器就是应用该原理而设计制作的。

颜料三原色的混合,亦称为减色混合,是将自然光中特定波长的光吸收,并反射剩下

的光而形成的。两色混合后,光度低于两色各自原来的光度,合色愈多,被吸收的光线愈多,就愈近于黑。所以,调配次数越多,纯度越差,越是失去它的单纯性和鲜明性。3种原色颜料的混合,在理论上应该为黑色。彩色印刷品是以黄、品红、青3种油墨加黑油墨印刷的,4色彩色印刷机的印刷就是一个典型的例证。在彩色照片的成像中,3层乳剂层分别为:底层为黄色、中层为品红、上层为青色。各品牌彩色喷墨打印机也都是以黄、品红、青加黑墨打印彩色图片的。

电视、电影是通过自身发光来合成颜色的,其合成法则被称为“加法原理”,三原色为红、绿、蓝,辅助色为“白”。印染、涂料则是通过吸收某些光线而形成颜色,因此其法则被称为“减法原理”,三原色为青、品、黄,辅助色为“黑”。

10. 彩色喷墨打印机工作原理

自然界中的色彩几乎都可以由选定的3种颜色以适当的比例混合得到,而且绝大多数的颜色也可以分解成特定的3种单色。通常,人们看到的彩色墨盒正是由几种纯净单一颜色组成,常见的3色墨盒打印机通常就是采用性质比较稳定的青色(C—Cyan)、品红色(M—Magenta)、黄色(Y—Yellow)来混合不同的颜色。而4色打印机,通常就加上一种黑色,用于纯黑色的打印。随着技术的发展,出现了6色墨盒,就是在原有的4色(CMYK)上再加上浅蓝绿色和浅红紫色。

假如将墨盒中的原色分别抽取不同的比例,再喷射到近似同一个点上,那么这个近似点便可以根据各原色不同的比例显示出不同的颜色,这就是彩色喷墨打印机原理。根据其喷墨方式的不同,可以分为热泡式喷墨打印机和压电式喷墨打印机两种。墨盒中的墨水经过压电式技术或者热喷式技术后,最终将不同的颜色喷射到一个尽可能小的点上,而大量这样的点便形成了不同的图案和图像,这一过程是一系列的繁杂程序。实际上,打印机喷头快速扫过打印纸时,它上面的无数喷嘴就会喷出无数的小墨滴,从而组成图像中的像素。打印喷头上一般都有48个或48个以上的独立喷嘴,每个喷嘴又能够喷出3种以上不同的颜色。一般来说,喷嘴越多,完成喷墨过程就越快,也就是打印速度越快。这些喷出来不同颜色的小墨滴落于同一点上,形成不同的复色。

在单色喷墨时代,这个点越小,图像将会越精细。业界通常用DPI来表示,意思即是在每英寸的范围内喷墨打印机可打印的点数。单色打印时DPI值越高打印效果越好。而彩色打印时情况比较复杂。通常打印质量的好坏要受DPI值和色彩调和能力的双重影响。其中,色彩调和能力是个非常重要的指标,传统的喷墨打印机,在打印彩色照片时,若遇到过渡色,就会在3种基本颜色的组合中选取一种接近的组合来打印,即使加上黑色,这种组合一般也不能超过16种,对彩色色阶的表达能力是难以令人满意的。

为了解决这个问题,早期的彩色喷墨打印机又采用了调整喷点疏密程度的方法来表达色阶。这就造成了一些分辨率低的打印品在近看的时候出现很多的小斑点。后来,人们想到了更好的办法,一方面通过提高打印密度(分辨率)来使打印出来的点变细,从而使

图变得更为精细。另一方面,都在色彩调和方面改进技术,常见的有增加色彩数量、改变喷出墨滴的大小、降低墨盒的基本色彩浓度等几种方法。其中,增加色彩数量来得最为行之有效。例如,6色墨盒,当打印机将6种不同颜色的墨滴喷到同一个点上,颜色组合最多可达64种,如果再结合不同大小的墨滴,便可能产生4096种不同的颜色。

11. 彩色激光打印机色彩合成原理

彩色激光打印机采用了青、品红、黄色、黑4色碳粉来实现全彩色打印,因此对于一页彩色内容中的彩色要经过CMYK调和实现,一页内容的打印要经过CMYK的4色碳粉各一次打印过程,所以目前大多数彩色激光打印机的彩色打印速度一般是黑白打印速度的四分之一。从理论上讲,彩色激光打印机要有4套与黑白激光打印机完全相同的机构来实现彩色打印过程。

最新彩色激光打印技术是所谓“一次成像”技术。这一技术的关键是需要把激光发光管做得足够小,在现有一个发光管的位置要放下对应于4种颜色的4个发光管。目前这一工艺的代价太高了,所以“一次成像”的彩色激光打印机价格昂贵,但这是未来的发展方向。

彩色激光打印机的彩色到底是如何合成的呢?下面以惠普的技术为例进行分析。ImageRet2400色彩分层技术是惠普所采用的技术。若确定打印的基本分辨率为600DPI,可以算出600DPI分辨率的图像,其像素之间的中心间距为 $42\mu\text{m}$ 。惠普使用直径为 $5\mu\text{m}$ 的Ultra Precise超精细碳粉,实际上可以实现2400DPI效果,2400DPI的分辨率时,像素之间的中心间距约为 $10\mu\text{m}$ 。也就是说,若最后彩色打印的结果是600DPI的像素分辨率,那么在一个像素点上,可以使用 16×16 个青、品、黄、黑的4种碳粉颗粒再加上空白来调制该像素点的颜色,由于人眼已无法分辨这些细微的颜色颗粒,所以人眼看到的是混色后的总效果。这种技术到底能提供多少种色彩,计算方法是排列组合中的一个经典的例题:有青、品、黄、黑、白5种颜色的无穷多个小球分装于5个坛子中,现在从中随意摸出16个小球放入一个空坛子中,一共会摸出多少种可能的结果?计算出的数值就是ImageRet2400技术理论上能实现的色彩数目,按厂家的说法是上百万种。若按现在的纳米材料的加工水平,碳微粒的直径达到5nm不成问题,那么现有的 $5\mu\text{m}$ 的碳粉颗粒还可以沿直径分开1000次,人们可以由此大胆地想像激光打印机理论上具有的色彩技术发展前景,当然颗粒细到一定的程度后对于人眼的识别来说已完全失去了继续细下去的意义。

12. AGP接口

随着图像应用软件的发展,在显卡和CPU以及主存中的数据交换量越来越大,而显卡的接口正是连接显卡和CPU的通道。现在的显卡都采用了AGP和AGP Pro接口。

AGP(accelerated graphics port)加速图形端口是在PCI图形接口的基础上发展而来的。随着大量三维应用程序的出现,原来传输速率为133MBps的PCI总线越来越不堪重负,由此拥有高带宽的AGP才得以浮出水面。这是一种与PCI总线迥然不同的图形接

口,它完全独立于 PCI 总线之外,直接把显卡与主板控制芯片联在一起,使得三维图形数据省略了越过 PCI 总线的过程,从而很好地解决了低带宽 PCI 接口造成的系统瓶颈问题。可以说,AGP 代替 PCI 成为新的图形端口是技术发展的必然。

1996 年 7 月,AGP 1.0 图形标准问世,分为 1X 和 2X 两种模式,数据传输带宽分别达到 266MBps 和 533MBps。这种图形接口规范是在 66MHz PCI 规范基础上经过扩充和加强而形成的,其工作频率为 66MHz,工作电压为 3.3V,在一段时间内基本满足了显示设备与系统交换数据的需要。

1998 年 5 月,AGP 2.0 规范正式发布,工作频率依然是 66MHz,但工作电压降低到了 1.5V,并且增加了 4X 模式,这样它的数据传输带宽达到 1066MBps。

最初的显示设备采用 PCI 总线接口,工作频率为 33MHz,数据宽度为 32 位,传输带宽为 $33 \times 32 \div 8 = 133\text{MBps}$ 。AGP 1X 的工作频率达到了 PCI 总线的两倍——66MHz,传输带宽理论上可达到 266MBps,一个时钟周期可以传送一次数据。AGP 2X 的工作频率同样为 66MHz,但是它使用一个时钟周期的上升沿和下降沿各传送一次数据,从而使一个工作周期先后被触发两次,即一个时钟周期可以传送两次数据,使传输带宽达到加倍的目的, $266\text{MBps} \times 2 = 532\text{MBps}$ 。AGP 4X 则规定一个时钟周期可以传送 4 次数据,这样在理论上它就可以达到 $266\text{MBps} \times 4 = 1064\text{MBps}$ 的带宽。

AGP 8X 也就是 AGP 2.2 标准协议,其重大的改进莫过于数据传输率的提升,采用新的更有效的信号安排,将传输率又提高了一倍,达到 2.128GBps。

7.4 教材习题解答

1. 外部设备有哪些主要功能? 可以分为哪些大类? 各类中有哪些典型设备?

解: 外部设备的主要功能有数据的输入、输出、成批存储以及对信息的加工处理等。外部设备可以分为五大类:输入输出设备、辅助存储器、终端设备、过程控制设备和脱机设备。其典型设备有键盘、打印机、磁盘、智能终端、数/模转换器和键盘—软盘数据站等。

2. 说明磁介质存储器的存储原理。

解: 磁介质存储器的存储过程是一种电磁转换的过程。在磁介质存储器中,信息是记录在一薄层磁性材料上的,这个薄层称为磁层。磁层与所附着的载体称为记录介质或记录媒体。磁头是磁记录设备的关键部件之一,是一种电磁转换元件,能把电脉冲表示的二进制代码转换成磁记录介质上的磁化状态,即电→磁转换;反过来,能把磁记录介质上的磁化状态转换成电脉冲,即磁→电转换。

3. 分别用 RZ、NRZ、NRZ-1、PE、FM、MFM 和 $M^2\text{FM}$ 制记录方式记录下述数据序列,画出写电流波形。

(1) 1101101110110

(2) 1010110011000

解：(1) 写电流波形如图 7-9(a)所示。

(2) 写电流波形如图 7-9(b)所示。

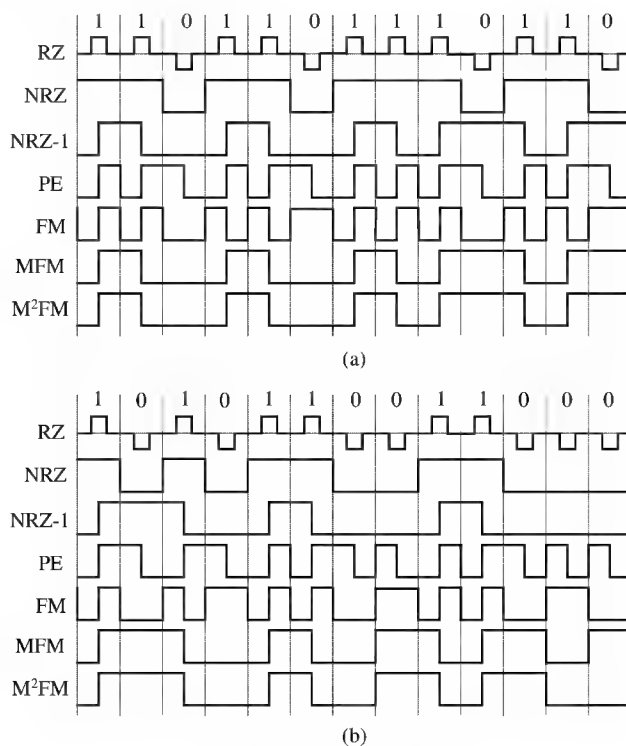


图 7 9 写电流波形

4. 若对磁介质存储器写入数据序列 10011,请画出不归零-1 制、调相制、调频制、改进的调频制等记录方式的写电流波形。

解：写电流波形如图 7-10 所示。

5. 主存储器与磁介质存储器在工作速度方面的指标有什么不同？为什么磁盘存储器采用两个以上的指标来说明其工作速度？

解：主存储器速度指标主要有存取速度和存取周期,而磁介质存储器速度指标为平均存取时间,这是因为磁介质存储器采用顺序存取或直接存取方式。磁盘存储器的平均存取时间至少应当包括平均寻道时间和平均等待时间两

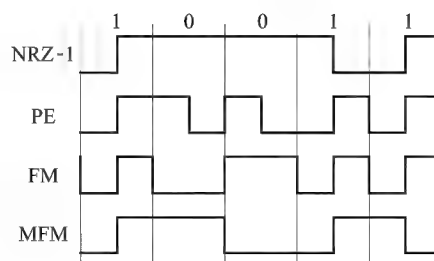


图 7 10 写电流波形

部分,因为磁盘存储器首先需要将磁头移动到指定的磁道上,然后将记录块旋转到磁头的下方才能进行读写。

6. 某磁盘组有六片磁盘,每片可有二个记录面,存储区域内径为 22cm,外径为 33cm,道密度 40 道/cm,位密度 400 b/cm,转速 2400 r/min。试问:

- (1) 共有多少个存储面可用?
- (2) 共有多少个圆柱面?
- (3) 整个磁盘组的总存储容量有多少?
- (4) 数据传送率是多少?
- (5) 如果某文件长度超过一个磁道的容量,应将它记录在同一存储面上还是记录在同一圆柱面上? 为什么?
- (6) 如果采用定长信息块记录格式,直接寻址的最小单位是什么? 寻址命令中如何表示磁盘地址?

解: (1) $6 \times 2 = 12$ (面), 共有 12 个存储面可用。

(2) $40 \times \frac{33-22}{2} = 220$ (道), 共有 220 个圆柱面。

(3) $12 \times 22\pi \times 400 \times 220 = 73 \times 10^6$ (位)。

(4) 数据传送率 = $\frac{22\pi \times 400}{60 \times 2400} = 1.1 \times 10^6$ (b/s) = 0.138×10^6 (B/s)。

(5) 记录在同一圆柱面上。因为这样安排存取速度快。

(6) 如果采用定长信息块记录格式,直接寻址的最小单位是扇区。磁盘地址为:驱动器号、圆柱面号、盘面号、扇区号。

7. 某磁盘存储器的转速为 3000 r/min,共有 4 个盘面,5 道/mm,每道记录信息 12 288B,最小磁道直径为 230mm,共有 275 道。试问:

- (1) 该磁盘存储器的存储容量是多少?
- (2) 最高位密度和最低位密度是多少?
- (3) 磁盘的数据传送率是多少?
- (4) 平均等待时间是多少?

解: (1) 磁盘存储器的容量 = $4 \times 275 \times 12\,288\text{B} = 13\,516\,800\text{B}$

(2) 最高位密度 D_1 = 每道信息量 ÷ 内圈圆周长 = $12\,288 \div (\pi \times \text{最小磁道直径}) \approx 17\text{B/mm}$ 。

最低位密度 D_2 = 每道信息量 ÷ 外圈圆周长 = $12\,288 \div (\pi \times \text{最大磁道直径}) \approx 11.5\text{B/mm}$ 。

(3) 磁盘数据传输率 $C = 50 \times 12\,288 = 614\,400\text{B/s}$ 。

$$(4) \text{ 平均等待时间} = \frac{1}{2r} = \frac{1}{2 \times 50} = 10\text{ms}。$$

8. 某磁盘组有效盘面 20 个, 每个盘面上有 800 个磁道。每个磁道上的有效记忆容量为 13 000B, 块间隔 235B, 旋转速度 3000 r/min。试问:

(1) 在该磁盘存储器中, 若以 1000B 为一个记录, 这样, 一个磁道能存放 10 个记录。若要存放 12 万个记录, 需要多少个圆柱面(一个记录不允许跨越多个磁道)?

(2) 这个磁盘存储器的平均等待时间是多少?

(3) 数据传送率是多少?

解: (1) 一个圆柱面可存放 200 个记录, 120000 个记录需要 600 个圆柱面。

(2) 平均等待时间为旋转半圈的时间, 10ms。

$$(3) \text{ 数据传送率} = \frac{13000}{20} = 650\text{KB/s}。$$

9. 某磁盘格式化为 24 个扇区和 20 条磁道。该盘能按需要选择顺时针或逆时针旋转, 旋转一圈的时间为 360ms, 读一块数据的时间为 1ms。该片上有 3 个文件: 文件 A 从磁道 6、扇区 1 开始占有 2 块; 文件 B 从磁道 2、扇区 5 开始占有 5 块; 文件 C 从磁道 5、扇区 3 开始占有 3 块。

试问: 该磁盘的平均等待时间为多少? 平均寻道时间是多少? 若磁头移动和磁盘转动不同时进行, 且磁头的初始位置在磁道 0、扇区 0, 按顺序 C、B、A 读出上述 3 个文件, 总的时间是多少? 在相同的初始位置情况下, 读出上述 3 个文件的最短时间是多少? 此时文件的读出次序应当怎样排列?

解: 平均等待时间为 180ms。磁盘分为 24 个扇区, 等待一个扇区的时间为 15ms。

平均寻道时间为磁头移动 10 条磁道的时间, 设移动一个磁道的时间为 n , 则平均寻道时间为 $10n$ 。

按顺序 C、B、A 读出上述 3 个文件, 总的时间包括:

总的寻道时间: 移动 5 道时间 + 移动 3 道时间 + 移动 4 道时间 = 移动 12 道时间 = $12n$

总的等待时间: $(3+1+9) \times 15 = 195\text{ms}$

总的读出数据时间: $(3+5+2) \times 1 = 10\text{ms}$

读出上述 3 个文件的最短时间包括:

总的寻道时间: 移动 2 道时间 + 移动 3 道时间 + 移动 1 道时间 = 移动 6 道时间 = $6n$

总的等待时间: $(5+7+5) \times 15 = 255\text{ms}$

总的读出数据时间不变。

此时文件的读出次序为 B、C、A。

10. 什么是光盘? 简述光盘的工作原理。

解: 相对于利用磁通变化和磁化电流进行读写的磁盘而言,用光学方式读写信息的圆盘称为光盘,以光盘为存储介质的存储器称为光盘存储器。

CD-ROM 光盘上有一条从内向外的由凹痕和平坦表面相互交替而组成的连续的螺旋形路径,当一束激光照射在盘面上,靠盘面上有无凹痕的不同反射率来读出程序和数据。

CD-R 光盘的写入是利用聚焦成 $1\mu\text{m}$ 左右的激光束的热能,使记录介质表面的形状发生永久性变化而完成的,所以只能写入一次,不能抹除和改写。

CD-RW 光盘是利用激光照射引起记录介质的可逆性物理变化来进行读写的,光盘上有一个相位变化刻录层,所以 CD-RW 光盘又称为相变光盘。

11. 键盘属于什么设备? 它有哪些类型? 如何消除键开关的抖动? 简述非编码键盘查询键位置码的过程。

解: 键盘是计算机系统不可缺少的输入设备。键盘可分为两大类型:编码键盘和非编码键盘。非编码键盘用较为简单的硬件和专门的键盘扫描程序来识别按键的位置。消除键开关抖动的方法分硬件和软件两种。硬件的方法是增设去抖电路;软件的方法是在键盘程序中加入延时子程序,以避开抖动时间。键盘扫描程序查询键位置码的过程为:

- ① 查询是否有键按下。
- ② 查询已按下键的位置。
- ③ 按行号和列号求键的位置码。

12. 说明针式打印和字模式打印有何不同? 各有什么优缺点?

解: 针式打印机利用若干根打印针组成的点阵来构成字符;字模式打印机将各种字符塑压或刻制在印字机构的表面上,印字机构如同印章一样,可将其上的字符在打印纸上印出。针式打印机以点阵图拼出所需字形,不需要固定字模,它组字非常灵活,可打印各种字符和图形、表格和汉字等,字形轮廓一般不如字模式清晰;字模式打印机打印的字迹清晰,但字模数量有限,组字不灵活,不能打印汉字和图形。

13. 什么是随机扫描? 什么是光栅扫描? 各有什么优缺点?

解: 扫描方式有两种:光栅扫描和随机扫描。

在光栅扫描方式中,电子束在水平和垂直同步信号的控制下有规律的扫描整个屏幕。这种方式的控制比较简单,画面质量较好且稳定,但对行扫描频率要求较高。

在随机扫描方式中,电子束能在屏幕上进行随机运动,其轨迹随显示内容变化而变化,只在需要显示字符和图形的地方扫描,而不必扫描全屏。这种方式显示速度快、画面清晰,尤其是线条的轮廓十分光滑,一般用于高清晰度的专用图形显示器中,但这种方式的控制比较复杂,而且只能用于字符和图形显示,不适于显示随机图像。

14. 什么是分辨率? 什么是灰度级? 它们各有什么作用?

解: 分辨率由每帧画面的像素数决定,而像素具有明暗和色彩属性。黑白图像的明

暗程度称为灰度,明暗变化的数量称为灰度级,分辨率和灰度级越高,显示的图像越清晰、逼真。

15.某字符显示器,采用 7×9 点阵方式,每行可显示60个字符,缓存容量至少为1260字节,并采用7位标准编码,试问:

(1)如改用 5×7 字符点阵,其缓存容量为多少?(设行距、字距不变——行距为5,字距为1。)

(2)如果最多可显示128种字符,上述两种显示方式各需多大容量的字符发生器ROM?

解:(1)因为显示器原来的缓存为1260B,每行可显示60个字符,据此可计算出显示器的字符行数: $1260\div 60=21(\text{行})$

因为,原字符窗口 $=8\times 14=(7+1)\times (9+5)$,现字符窗口 $=6\times 12=(5+1)\times (7+5)$ 。

所以,现显示器每行可显示80个字符,显示器可显示的字符行数为24行。

故缓存的容量为 $80\times 24=1920\text{B}$ 。

(2)ROM中为行点阵码

7×9 点阵方式: $128\times 9\times 7=1152\times 7(\text{位})=1152(\text{字节})$

5×7 点阵方式: $128\times 7\times 5=896\times 5(\text{位})=896(\text{字节})$

注:为存储方便,每个行点阵码占用一个字节。

16.某CRT显示器可显示64种ASCII字符,每帧可显示64列 \times 25行,每个字符点阵为 7×8 ,即横向7点,字间间隔1点,纵向8点,排间间隔6点,场频50Hz,采用逐行扫描方式。试问:

(1)缓存容量有多大?

(2)字符发生器(ROM)容量有多大?

(3)缓存中存放的是字符的ASCII码还是字符的点阵信息?

(4)缓存地址与屏幕显示位置如何对应?

(5)设置哪些计数器以控制缓存访问与屏幕扫描之间的同步?它们的分频关系如何?

解:(1)缓存容量: $64\times 25=1.6\text{KB}$ (不考虑显示属性), $64\times 25\times 2=3.2\text{KB}$ (考虑显示属性)。

(2)字符发生器(ROM)容量 $=64\times 8=512\text{B}$ 。

(3)缓存中存放的是字符的ASCII码。

(4)屏幕显示位置自左至右,从上到下,相应地缓存地址由低到高,每个地址码对应一个字符显示位置。设字符在屏幕上的位置坐标为 (X,Y) ,即行地址为 X ,列地址为 Y ,则缓存地址 $=X\times 80+Y$ (未考虑显示属性)。

(5)设置4个计数器以控制缓存访问与屏幕扫描之间的同步。

它们的分频关系是:

点计数器: 8 分频(包括横向 7 点和字间间隔 1 点)。

字计数器: 79 分频(包括一行显示 64 个字符和水平回扫折合的字符数)。

行计数器: 14 分频(包括纵向 8 点,排间间隔 6 点)。

排计数器: 26 分频(包括显示 25 排字符和垂直回扫折合的字符排数)。

17. 某 CRT 字符显示器,每帧可显示 80 列 \times 20 行,每个字符是 7 \times 9 点阵,字符窗口 9 \times 14,场频为 50Hz。试问:

(1) 缓存采用什么存储器,其中存放的内容是什么? 容量应为多大?

(2) 缓存地址如何安排? 若在 243 号单元存放的内容要显示出来,其屏幕上 X 和 Y 的坐标应是多少?

(3) 字符点阵存放在何处? 如何读出显示?

(4) 计算出主振频率以及点计数器、字计数器、行计数器、排计数器的分频频率。

解: (1) 缓存采用随机存储器,其中存放的内容是字符的 ASCII 码,容量至少为 1600 字节(不含显示属性)。

(2) 屏幕上最多可显示 1600 个字符,缓存地址与屏幕显示位置的排号和列号具有对应关系。若要将缓存 243 号单元存放的内容显示出来,其屏幕上 X 和 Y 的坐标均为 3(从 0 开始计),即在屏幕的第 4 行第 4 列上有字符显示。

(3) 字符点阵存放在字库中,根据字符的 ASCII 码逐行读出点阵显示。

(4) 主振频率 $=50 \times 21 \times 14 \times 98 \times 9 \approx 12.97\text{MHz}$ 。

点计数器: 9 分频。

字计数器: $(80+18)$ 分频。

行计数器: 14 分频。

排计数器: $(20+1)$ 分频。

18. 若用 CRT 作图形显示器,其分辨率为 640 \times 200,沿横向每 8 点的信息存放在缓存中,场频为 60Hz。试问:

(1) 缓存的基本容量是多少?

(2) 地址如何安排?

(3) 点计数器、字节计数器、行计数器各为多少分频?

(4) 它和字符显示器有哪些不同?

解: (1) 缓存的基本容量是 16000 字节(不考虑灰度级)。

(2) 缓存地址为行号 \times 80+列号。

(3) 点计数器: 8 分频。

字节计数器: $(80+L)$ 分频,其中 L 次计数,作为行线逆程回扫折合的字节数。

行计数器: $(200+M)$ 分频,其中 M 次计数,作为场逆程回扫折合的行数。

(4) 图形显示器和字符显示器不同在于:图形显示器需将每个像素的信息都存放在 VRAM 中,而字符显示器只需将要显示的 ASCII 码存放在 VRAM 中,字符的点阵来自字符发生器 ROM。

19. 某字符显示器分辨率为 40 列 \times 25 行,字符点阵 5 \times 7,横向间隔 2 点,排间间隔 4 点,问:缓存 VRAM 容量至少应多大?应设置哪几级同步计数器?它们的分频关系如何?若要求场频 60Hz,则点频应为多少?何时访问一次 VRAM?地址如何确定?

解: 缓存 VRAM 容量至少 1000 字节。

设置 4 级同步计数器(点计数器、字计数器、行计数器、排计数器),它们的分频分别是点计数器 7 分频、字计数器(40+ L)分频、行计数器(7+4)分频、排计数器(25+ M)分频。其中, L 是水平回扫折合的字符数, M 是垂直回扫折合的字符排数。假设 $L=9$, $M=1$,有:

$$\text{点频} = 60 \times 26 \times 11 \times 49 \times 7 \approx 5.89 (\text{MHz})$$

每隔 $1.189\mu\text{s}$ (字符脉冲频率的倒数)访问一次 VRAM,地址由字计数器和排计数器共同提供,其中,字计数器提供低位地址,行计数器提供高位地址。

20. 某图形显示器的分辨率为 800 \times 600,若作单色显示且不要求灰度等级,则 VRAM 容量至少应多大?应设置哪几级同步计数器?它们的分频关系如何?若要求场频 60Hz,则点频应为多少?何时访问一次 VRAM?地址如何确定?

解: VRAM 容量至少应为 60 000 字节。

设置 3 级同步计数器(点计数器、字节计数器、行计数器),它们的分频关系分别是点计数器 8 分频、字节计数器(100+ L)分频、行计数器(600+ M)分频。其中, L 是行线逆程回扫折合的字节数, M 是场逆程回扫折合的行数。假设 $L=23$, $M=10$,有:

$$\text{点频} = 60 \times 610 \times 123 \times 8 \approx 36 (\text{MHz})$$

每隔 $0.22\mu\text{s}$ 访问一次 VRAM,地址由字节计数器和行计数器共同提供,其中字节计数器提供列号,行计数器提供行号。

21. 水平扫描频率(行频)的单位为 kHz,垂直扫描频率(场频)的单位为 Hz,两者为何相差一千倍?

解: 行频又称水平扫描频率,是电子枪每秒在屏幕上扫描过的水平线条数,以 kHz 为单位。场频又称垂直扫描频率,是每秒钟屏幕重复绘制显示画面的次数,以 Hz 为单位。因为每一场有近千条水平扫描线,所以行频与场频要相差近一千倍。

第 8 章

输入输出系统

8.1 基本内容要求

计算机的输入输出系统是整个计算机系统中最具有多样性和复杂性的部分,本章首先讨论主机与外设之间的连接问题,接着重点介绍程序查询方式、程序中断方式、DMA 方式、通道方式,最后介绍总线技术和接口标准。

学习要求

- ◆ 了解接口的基本组成
- ◆ 理解输入输出接口和端口概念的不同
- ◆ 了解接口的类型
- ◆ 了解外设的识别与端口寻址
- ◆ 了解各种输入输出信息传送控制方式的特点和适用范围
- ◆ 了解程序查询方式的特点和 workflow
- ◆ 理解程序中断的基本概念
- ◆ 理解程序中断与调用子程序的区别
- ◆ 了解程序中断的基本类型
- ◆ 掌握 CPU 响应中断的 3 个条件
- ◆ 掌握中断隐指令的特点以及中断隐指令完成的 3 个操作
- ◆ 理解进入中断服务程序的方法(向量地址的形成)
- ◆ 了解中断现场的保护和恢复方法
- ◆ 理解中断允许触发器的作用以及开、关中断的时机
- ◆ 掌握中断屏蔽的概念,通过改变中断屏蔽字实现中断升级
- ◆ 掌握 DMA 方式的特点
- ◆ 理解 DMA 方式和程序中断方式的区别
- ◆ 了解 DMA 接口的组成

- ◆ 掌握 DMA 传送方法和 DMA 传送过程
- ◆ 理解通道控制方式与 DMA 方式的区别
- ◆ 了解通道的类型与结构
- ◆ 了解通道工作过程
- ◆ 了解总线通信控制方式的特点
- ◆ 理解 3 种总线判优和仲裁方式的区别

8.2 误点疑点解惑

1. 接口与接口中的寄存器

输入输出接口是主机和外设之间的交接界面,通过接口可以实现主机和外设之间的信息交换。接口的基本组成以及与主机、外设间的连接如图 8-1 所示。接口中要分别传送数据信息、控制信息和状态信息,控制信息是指 CPU 向设备发出的命令信号,状态信息是指外设和接口向主机报告的信息。在许多接口电路中将数据信息、控制信息和状态信息都看作广义的数据信息,其中控制信息视为输出数据,状态信息视为输入数据,均通过数据线与主机进行交换。

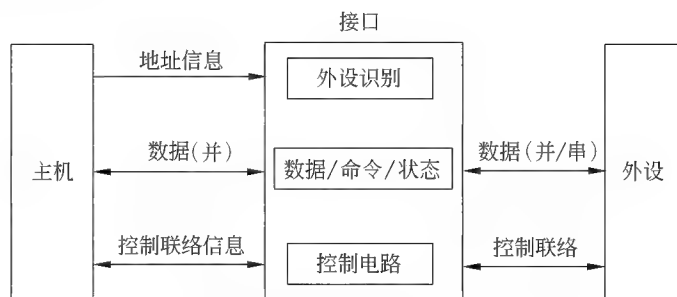


图 8-1 接口与主机、外设间的连接示意图

接口电路与主机一侧连接的数据线有多根,并行传送多位数据,与设备一侧连接的数据线的数量取决于外设的类型,若是串行外设,则只有一根数据线,若是并行外设,则有多根数据线。

在接口电路中设置有能供 CPU 直接访问的寄存器,这些寄存器被称为端口,不同的端口被赋予不同的地址。通常,一个接口中包含有数据端口、命令端口和状态端口。存放数据信息的寄存器称为数据端口,存放控制命令的寄存器称为命令端口,存放状态信息的寄存器称为状态端口。CPU 通过输入指令可以从有关端口中读取信息,通过输出指令可以把信息写入有关端口。有的端口只能写或只能读,有的端口既可以读又可以写。例如,对状态端口只能读,在 80x86 中用输入指令(IN AL, 状态口地址)可将外设的状态标志送

到 CPU 中去;对命令端口只能写,在 80x86 中用输出指令(OUT 控制口地址,AL)可将 CPU 的各种控制命令发送给外设。为了节省硬件,在某些接口电路中,状态信息和控制信息可以共用一个寄存器,称之为设备的控制/状态寄存器。

2. 程序查询方式传送举例

采用程序查询方式进行输入输出时,主机与外设之间处于串行工作状态。就是说,外设工作时,主机什么也不能干,只是不停地查询该外设的工作状态,直到查询到外设已传送完一个数据时,则用很短的时间将这个数据取走。显然,采用这种工作方式,CPU 只能与低速的外设交换信息。

下面以计算机早期使用的一种输入设备——光电输入机为例来说明程序查询方式的操作过程。假定某光电机以 1KB/s 的速度运行,即每输入一个字节需要 1ms 的时间。现有一批数据已记录在穿孔纸带上,等待输入到主存的一片区域中。

CPU 首先要将本次输入的字节数和主存首地址置入相应的寄存器,并启动光电机工作,然后不断地查询光电机的工作状态,直到光电机走完一排孔,即输入完一个字节,则用很短的时间将该字节写入主存中,并修改计数值和主存地址。重复上述过程,直至整个数据块输入完毕,操作过程可用流程图来描述,假设查询过程(图 8-2 中虚线框内)需要两条指令,指令周期为 $1\mu\text{s}$,而光电机要 1ms 才能准备好一个数据,所以,在整个数据块输入过程中,CPU 约有 99% 以上的时间处于查询等待状态而不能进行任何其他操作,显然对于 CPU 来说这是一种效率很低的输入方式。但这种方式控制简单,只需要很简单的硬件支持,有几个相关的寄存器和计数器就足够了。

3. 中断系统的功能

现代计算机无论是巨型机、大型机、小型机还是微型机无不具有中断处理的能力。纵观各类计算机的中断系统,其功能可概括为以下几个方面:

① 实现主机和外设的并行工作。采用中断方式在主机和外设之间相互交换信息,以完成输入输出功能,由于主机和外设并行工作,从而大大提高了计算机的工作效率。

② 及时发现和处理机器中的软、硬件故障。在计算机工作时,当运行程序发生故障或硬件发生故障时,机器中断系统可以中断正在处理的程序而自动进入故障处理程序,避免了某些偶然事故引起的计算故障或停

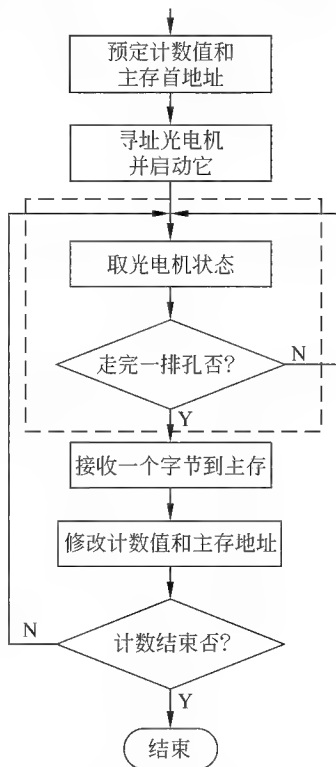


图 8-2 光电机输入操作流程

机,提高了机器的可靠性。

③ 进行实时处理。实时处理是指在某个时间或现象出现的实际时间内及时地进行处理,而不是积压起来再进行批处理。例如,某个计算机过程控制系统中,在生产过程中采集到随机出现的压力过大、温度过高等信息后,应及时地输入计算机,并以最快速度立即响应而不允许延迟。

④ 实现人机对话。某些程序执行后,需要操作人员由控制台、打字机或某种终端设备提供信息,进行控制。可利用中断系统提供人工干预的途径,以实现人机对话功能。

⑤ 实现多道程序的并行执行。在多用户计算机系统中,可通过中断系统实现多道程序之间的调度。多道程序运行时,对每道程序分配一个固定的时间片,利用时钟定时发中断请求,进行程序切换。

4. 程序中断方式不适合高速外设数据传送的原因

采用程序中断方式进行数据传送可以提高主机效率,但这种方式并不适合于高速外设和主存储器之间的数据传送。若用于高速外设中,会使主机处于频繁的中断和返回过程中,从而加重了与中断有关的额外负担(即保护旧现场,恢复新现场),降低了主机的性能,还有发生丢失数据的可能。例如,磁盘平均速度为 $100\,000\text{B/s}$,即用 $10\mu\text{s}$ 传送一个字节;若 CPU 的指令周期为 $2\mu\text{s}$,中断服务程序为 10 条指令,交换一次数据需要 CPU 花费 $20\mu\text{s}$ 。其结果显然是,第一个数据还没有取走,第二个数据便将第一个数据冲掉,致使数据丢失。所以,对高速外设用程序中断方式是不适合的。

5. 中断响应阶段完成的任务

中断源发出中断请求后,CPU 中止现行程序的执行,转去为某个中断源服务的过程称为中断响应。中断响应阶段是中断全过程中至关重要的一个阶段,同时,由于不同计算机的中断系统具有一定的差异,不同计算机的中断响应阶段完成的任务也有所不同,所以这部分将成为教学中的一个难点。

CPU 响应中断要满足 3 个条件:①中断源有中断请求;②CPU 允许接受中断请求;③一条指令执行完毕。第①个条件是显而易见的,无须多说,而第②、③个条件则需要仔细讨论。CPU 内部有一个中断允许触发器(注意并非每个中断源都有一个),以此来确定 CPU 的现行程序是否可以被中断。当中断允许触发器=1 时,CPU 处于开放状态,允许中断;当中断允许触发器=0 时,CPU 处于关闭状态,禁止中断。中断允许触发器由开中断指令来置位,由关中断指令或硬件自动使其复位。CPU 响应中断的时间是在一条指令(这条指令不能是停机指令)执行完毕,且没有优先权更高的请求(如电源失效或 DMA 请求)时,随后 CPU 进入中断周期。之所以必须要等到一条指令执行完毕,是因为响应中断意味着处理机将从一个程序(现行程序)切换到另一个程序(中断服务程序),而程序是由一条条的指令组成的,如果不在指令执行完毕时进行程序的切换,中断返回时将无法保证原来的程序能继续执行。

为了使得切换前后的程序都能正确地运行,在中断响应阶段需要将 CPU 的关键性硬件状态保存起来。这些状态主要有两类:一类是表示程序进程的`程序状态字 PSW`和标志进程轨迹的`程序计数器 PC(断点)`,另一类是一些工作寄存器(如通用寄存器等),它们保存着程序执行的现行值,称之为中断现场。`PSW`和`PC`的内容必须在程序被中止时就加以很好地保护,以便在恢复时程序能正确地沿断点继续执行,否则`PC`和`PSW`的内容在中断响应时将被中断服务程序的入口地址和中断服务程序的状态字冲掉,所以往往在中断周期中由硬件来完成它们的保存。而工作寄存器的内容在中断响应时不会被破坏,因此可以在中断服务程序里由软件把它转移到其他安全地方去。

在中断周期,CPU 执行一条中断隐指令。中断隐指令由硬件在中断响应时产生,它并不是指令系统中的一条真正的指令,本身没有操作码,也不会出现在程序中。中断隐指令主要完成 3 个操作:①保存断点;②关闭中断允许触发器;③找出中断服务程序的入口地址。

在中断周期中必须关闭中断允许触发器的原因是,保证用软件来保护现行程序的中断现场期间不允许被新的、更高级的中断请求所打断。并不是所有的计算机都一定在中断隐指令中由硬件来关闭中断允许触发器,也有些计算机的这一操作是在中断服务程序中的保护现场之前由关中断指令来实现的。

在公共请求线的中断系统中,CPU 响应中断后,必须识别出发出中断请求的中断源,才能找到中断服务程序的入口地址。最简单的方法是采用软件查询的方法,软件查询中断源是与中断判优结合在一起的。在中断周期中给出的是公共服务程序的入口地址,公共服务程序是一段查询程序,当查询到中断请求的发出者,也就是找到了中断源,程序可以立即转入对应的中断服务程序中去,为相应的中断源服务。查询程序根据查询顺序先后确定优先级,改变查询的先后次序就可以改变优先级。这种方法节省了硬件开销,但增加了查询时间。另一种识别中断源的方法是串行优先链,在 CPU 接收到中断请求信号`INTR`之后,回复一个中断响应信号`INTA`,通过硬件排队电路查出优先级最高的中断源,CPU 再找出该中断源的中断服务程序入口地址。这种方法具有很高的响应速度,其缺点是中断源的优先级被硬件排队电路固定死,使中断系统不够灵活。

6. 中断服务程序入口地址的获取方式

向量中断或非向量中断获取中断服务程序入口地址的方式是不同的。向量方式通过硬件方式确定中断源,产生对应于中断源的向量地址,可以快速直接转向对应的中断服务程序;非向量中断则是通过软件方式确定中断源,再分支进入相应的中断服务程序。

现代计算机基本上都具有向量中断功能,其具体实现方法有多种。比如,在`80x86`系统中,由中断类型码 $\times 4$,形成向量地址;在`Z-80`系统中,中断源可以送出一种复位指令`RST n`,再转换成向量地址。又如,在具有多根请求线的系统中,可由请求线编码产生各中断源的向量地址。再如,在菊花链结构中,经硬件查询找到批准的中断源,该中断源通

过总线向 CPU 送出其向量地址。在有些系统中,CPU 内有一个中断向量寄存器,存放向量地址的高位部分,中断源产生向量地址的低位部分,两者拼接形成完整的向量地址。

CPU 在响应非向量中断时只产生一个固定的地址,由此读取中断查询程序的入口地址,进而转向查询程序,通过软件查询,确定被优先批准的中断源,然后分支进入相应的中断服务程序。在 DTS-130 机中,CPU 响应中断时,从主存的 1 号单元中读出查询程序的入口地址,然后转向查询程序,通过执行查询程序,按优先顺序逐个查询各中断源。若某中断源提出请求,则转向相应的中断服务程序;若未提出请求,则继续往下询问。查询程序是为所有中断请求服务的,又称为中断总服务程序,它的任务仅仅是判定优先级最高的中断源,从而转向实质性处理的服务程序。查询程序本身可以存放在任何主存空间,但它的入口地址被写入一个固定的单元,各个中断服务程序的入口地址则被写进查询程序之中。查询方式可以是软件查询,也可以先通过硬件取回被批准中断源的设备码,再通过软件判别。

7. 程序中断方式中容易混淆的几个问题

(1) 中断系统的概念

中断系统是计算机实现中断功能的软、硬件总称。一般在 CPU 中配置中断机构,在外设接口中配置中断控制器,在软件上设计相应的中断服务程序。

通常,每个中断源都有自己的中断请求触发器和中断屏蔽触发器,多个中断源的这两个触发器可以组成多位的中断请求寄存器和中断屏蔽寄存器;而 CPU 中只有一个中断允许触发器,由它来控制是否允许中断。

(2) 禁止中断与屏蔽中断

禁止中断与屏蔽中断是两个完全不相关的概念,但学生在学习的过程中很容易将两者混为一谈。

禁止中断是指 CPU 中的中断允许触发器被置“0”,此时中断关闭(关中断),所有中断源的中断请求都不能得到响应。与禁止中断相对应的是允许中断,即中断允许触发器被置“1”,此时中断允许(开中断),来自中断源的中断请求可以得到响应。适时的开、关中断,将使 CPU 能正确地进行程序切换。例如,为了保证多重中断时保护和恢复现场工作的完整性,在保护和恢复现场之前必须关中断,在保护和恢复现场之后必须开中断。

屏蔽中断是指某个中断源的中断屏蔽触发器被置“1”,此时对应的中断源不能请求中断服务。与屏蔽中断相对应的是开放中断,即中断屏蔽触发器被置“0”,此时对应的中断源可以请求中断服务。各个中断源的中断屏蔽位组合起来形成一个中断屏蔽码,利用修改屏蔽码可以将某些中断源的中断请求暂时屏蔽起来,以此来改变 CPU 为中断源服务的先后次序,达到在有多个中断源同时请求中断时,先为较低级中断源服务,然后再为较高级中断源服务的目的。

综上所述,禁止中断是对全部中断源的中断请求均加以禁止,而屏蔽中断只是将部分

中断源的中断请求加以屏蔽。

(3) 中断响应次序和中断处理次序

中断处理次序和中断响应次序是两个不同的概念,中断响应次序是由硬件排队电路决定的,一旦排队电路设计完成将无法改变。但是,中断处理次序是可以由中断屏蔽码来改变的,所以把中断屏蔽码看成软排队器。正常情况下,中断处理次序就等于中断响应次序,但如果由程序员改变了中断屏蔽码,中断处理次序就不同于中断响应次序了。

8. DMA 控制器的控制过程

采用程序查询或程序中断方式,外设与主机交换信息是完全在 CPU 的控制下进行的,输入输出操作给 CPU 增加了很大的额外开销。采用 DMA 方式,是在主存与外设之间建立一条“直接数据通路”,在不需 CPU 干预也不需要软件介入的情况下在两者之间进行的高速数据传送方式。需要提醒学生注意的是,并不是真正在主存与外设间建立一条物理的直接数据通路,而是利用系统总线在主存与外设间建立一条逻辑的直接数据通路。另外,有些学生在答题中可能会将“主存和外设”之间误回答为“主机和外设”、“CPU 和外设”之间,所以,一定要强调在传送的过程中是不需要 CPU 干预的。

正常情况下,CPU 拥有对系统总线的控制权,在采用 DMA 方式时,CPU 要放弃对系统总线的控制权,而将其赋予 DMA 控制器。具体操作过程如下:

当某个外设端口向 DMA 控制器发出 DMA 请求信号(DREQ)时,由 DMA 控制器向 CPU 发出总线请求信号(HRQ),CPU 接收到这一请求后,将在正在执行的指令的当前机器周期结束时响应这一请求,并向 DMA 控制器回送一个总线响应信号(HLDA),表示 CPU 从现在开始放弃对系统总线的使用权,于是 DMA 控制器进入主控状态,向发出请求的外设端口回送 DMA 响应信号(DACK),并在主存和外设端口之间建立一条直接传送数据的通路,完成输入输出操作功能,其控制过程如图 8-3 所示。

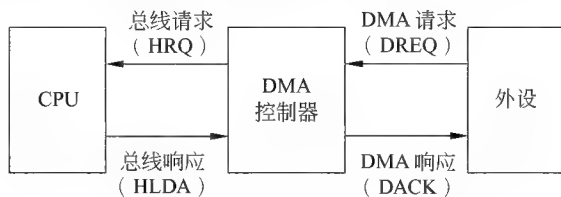


图 8 3 DMA 控制器的控制过程

DMA 控制器是一个特殊的接口,当 CPU 获得系统总线控制权时,它处于从属状态,接收 CPU 对它的控制(如 DMA 传送前的预处理);当 CPU 放弃对系统总线的控制权时,它处于主控状态,接管对系统总线的控制权,完成整个数据传送过程的控制;待数据传送完毕,CPU 将收回对系统总线的控制权,DMA 控制器重新回到从属状态。通常将获得总线控制权的设备称为主设备,与之进行信息交换的对象称为从设备,所以 DMA 控制器有

时是主设备,有时是从设备。

9. 单字传送与成组连续传送

每提出一次 DMA 请求将占用多少个总线周期? 是单字传送还是成组连续传送? 合理的安排 CPU 访存与 DMA 传送中的访存,是计算机系统设计时应该考虑的问题。采用 DMA 方式是为了实现一次批量传送,如从磁盘中读出一个文件,但具体实施上有两类方案。

(1) 单字传送

每次 DMA 请求获得批准后,CPU 让出一个周期的总线控制权,由 DMA 控制器控制系统总线,以 DMA 方式传送一个字节或一个字。然后,DMA 控制器将系统总线控制权交回 CPU,重新判断下一个周期的总线控制权归属。这种方式称为单字传送方式,也就是周期窃取方式。每次 DMA 请求从 CPU 控制中挪用一個总线周期(DMA 周期),用于 DMA 传送。

当主存工作速度高出外设较多时,采用单字传送方式可以提高主存利用率,对 CPU 程序执行的影响较小。因此,高速主机系统常采用这种方式,这是因为在 DMA 传送数据尚未准备好(例如,尚未从磁盘中读到新的数据)时,CPU 可用系统总线访问主存。根据主存读写周期与磁盘的数据传送率,可以算出主存操作时间的分配情况:有多少时间需用于 DMA 传送(被挪用),有多少时间可用于 CPU 访存,这在一定程度上反映了系统的处理效率。由于访存冲突,每次 DMA 传送会对 CPU 正常执行程序带来一定的影响,但由于主存速度较高,因而影响并不严重。

(2) 成组连续传送

每次 DMA 请求获得批准后,DMA 控制器掌管总线控制权,连续占用若干个总线周期,进行成组连续的批量数据传送,直到批量传送结束,才将总线控制权交还 CPU。在传送期间 CPU 处于保持状态,停止访问主存,因此也就无法执行程序。

当外设的数据传送率接近于主存工作速度时,常采用成组连续传送方式。这种方式可以减少系统总线控制权的交换次数,有利于提高输入输出速度。由于系统必须优先满足 DMA 高速传送,如果 DMA 传送的速度接近于主存速度,则每个总线周期结束时将总线控制权交回 CPU 就没有多大的意义了。对单用户个人计算机,一旦启动调用磁盘,CPU 就等待这次调用结束才恢复执行程序,因此也可等到批量传送结束才收回总线控制权。对于高速计算机,常用多道程序工作方式,且主存速度超出外设速度很多,如果采用成组连续传送方式,就会影响主机的利用率。

10. 周期挪用法的特点

周期挪用法是一种对 CPU 工作影响最小的 DMA 传送方法。基本做法是:CPU 在工作过程中,一旦查询到有 DMA 控制器产生的总线请求信号时,则暂停本身的操作,将一个周期让给 DMA 控制器,由它控制通过总线传送一个字节或一个字,然后 CPU 继续

进行自己的操作,等待下一个总线请求的到来,重复上述过程,直到总线请求信号发完为止。可以看出这种方式基本上不影响 CPU 的工作,只是在需要时将一个周期挪用给 DMA 控制器,或者说,DMA 控制器在需要时从 CPU 窃取一个周期。利用周期挪用法进行 DMA 传送的过程如图 8-4 所示。

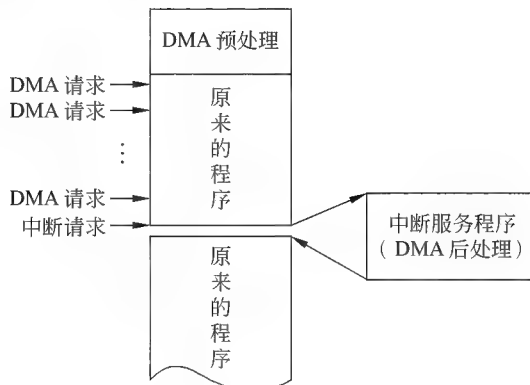


图 8-4 DMA 传送的过程

由于从 CPU 窃取周期的目的是为了完成与主存的信息交换,所以这里所说的周期当然是指存取周期。在许多计算机中,为了简化时序控制,往往令存取周期就等于机器周期;在微型计算机中,主存是连接在系统总线上的,因此存取周期也就是总线周期。此时,存取周期与机器周期、总线周期等概念的含义是相同的,但要注意这一切都是有前提的。

周期挪用相当于根据需要在 CPU 正常的工作期间插入一个存储周期,用于 DMA 传送。当 CPU 的访存速度高出外设较多时,采用周期挪用法对 CPU 的影响很小。

11. 采用周期挪用法的数据传送过程

下面以数据输入为例,讨论周期挪用法的数据传送过程。

(1) 由主程序启动设备,从设备读入一个字到 DMA 控制器的数据缓冲寄存器中。此时设备控制器完成信号将 DMA 控制器中的 DMA 请求触发器置“1”,表示设备已完成一个数据传送工作。DMA 控制器向 CPU 发出总线请求,申请存取周期。

(2) CPU 响应总线请求并在 CPU 的一个存储周期结束后,放弃对系统总线的控制权,DMA 控制器立即占用下一个存取周期(DMA 周期)进行操作,此时 CPU 现场冻结。图 8-5 为 DMA 运动轨迹。在 DMA 周期中 DMA 控制器获得总线的控制权,执行以下 3 个操作:

- ① 将存储器数据区首地址送主存地址计数器;
- ② 将输入数据送数据缓冲寄存器;
- ③ 发送写存储器命令。

当 DMA 周期结束后以清除信号送 DMA 控制器。

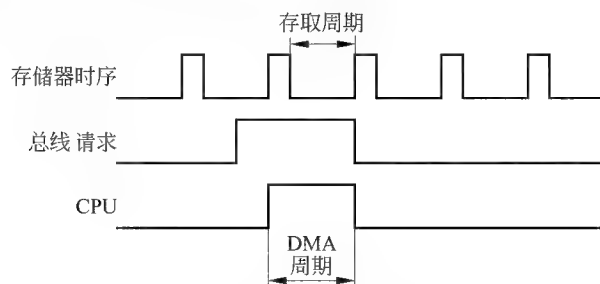


图 8 5 DMA 运动轨迹

(3) 清除信号在 DMA 控制器中执行 3 个操作：

- ① 传送长度计数器减 1；
 - ② 主存地址计数器的内容加 1 指向存储器数据区的下一个地址；
 - ③ 将 DMA 请求触发器置“0”，以表示本次 DMA 结束。
- (4) 高速设备只需要启动一次，以后连续不断读出，直至完成全部数据的传送。
- (5) 数据全部读出并交换完毕后（传送长度计数器为 0），DMA 控制器发中断请求，请求 CPU 进行结束处理。

12.3 种不同类型通道的比较

通道有 3 种不同的类型：字节多路通道、选择通道和数组多路通道。

字节多路通道适用于连接大量字符类低速设备，通道的数据宽度（每次传送的数据量）为单字节，以字节交叉方式轮流地为多台外部设备服务。选择通道和数组多路通道都适用于连接高速外设，但前者的数据宽度是不定长的数据块，后者的数据宽度是定长的数据块。3 种类型通道的简要比较如表 8-1 所示。

表 8 1 3 种类型通道的区别

通道类型 性能	字节多路	选 择	数组多路
数据宽度	单字节	不定长块	定长块
适用范围	大量低速设备	优先级高的高速设备	大量高速设备
工作方式	字节交叉	独占通道	成组交叉
共享性	分时共享	独占	分时共享
选择设备次数	多次	一次	多次

选择通道和数组多路通道在对高速外设服务方式是完全不同的，一条选择通道在物理上可连接多个外设，但是在逻辑上一条选择通道只能连接一台外设，这就是说，选择通

道任何时候只能为一台外设所独占,或者说,不管一条选择通道上连接了多少台外设,任何时候只能在一台外设与主存之间建立数据传送的通路,只有等这台设备从寻址到一个数据块传送结束,才有可能去为别的外设服务。而一条数组多路通道无论物理上还是逻辑上都可以连接多台外设,数组多路通道是以数据块交叉的方式同时为多台高速外设服务。具体地说就是利用为某一台外设寻址的时间去为另一台外设传送数据,从时序上来看可用图 8-6 来描述。

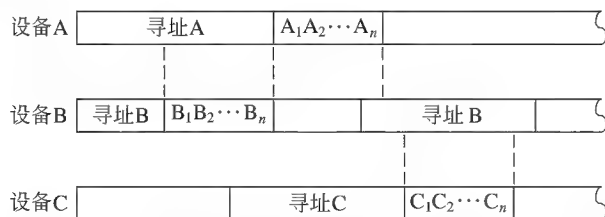


图 8 6 数组多路通道的传输过程

与选择通道比较,可认为数组多路通道上设置有多个子通道,各个子通道独立执行自己的通道程序为某个高速外设服务,从宏观上来看,数组多路通道在并行地为多个外设服务,显然,数组多路通道比选择通道具有更高的效率和利用率,只是在控制上更复杂。

13. 通道操作的全过程

CPU 需要进行输入输出操作时,在用户程序中使用访管指令迫使 CPU 由用户程序(目态)进入管理程序(管态),通过执行相应的访管子程序,根据访管指令给定的参数编写程序(通道程序)写入主存的一片区域中,并将其首地址置入通道地址字(CAW)中,然后,便可启动该通道开始工作,CPU 返回用户程序的断点 $K+n$ 继续工作。从此时开始,CPU 与通道处于并行工作状态。通道从 CAW 中获得通道程序的入口地址,逐条取出通道指令并执行它,待通道程序执行完毕可向 CPU 发出中断请求,CPU 响应该中断请求,再次进入管理程序进行结束处理,本次输入输出操作完成。上述过程如图 8-7 所示。

从图 8-7 中可以看出,整个输入输出操作的过程是在通道控制下完成的,而通道的控制是通过执行通道程序实现的。

在通道操作过程中,CPU 只需要进行两次干预,第一次是在访管指令的要求下,执行访管子程序,编写好通道程序并写入主存的一片区域中,再将通道程序在主存中的首地址置入 CAW 中,并启动该通道开始执行通道程序;第二次是待通道程序执行完毕,输入输出操作完成后,响应通道的中断请求进入结束中断服务程序,完成对本次输入输出操作的结束处理过程。

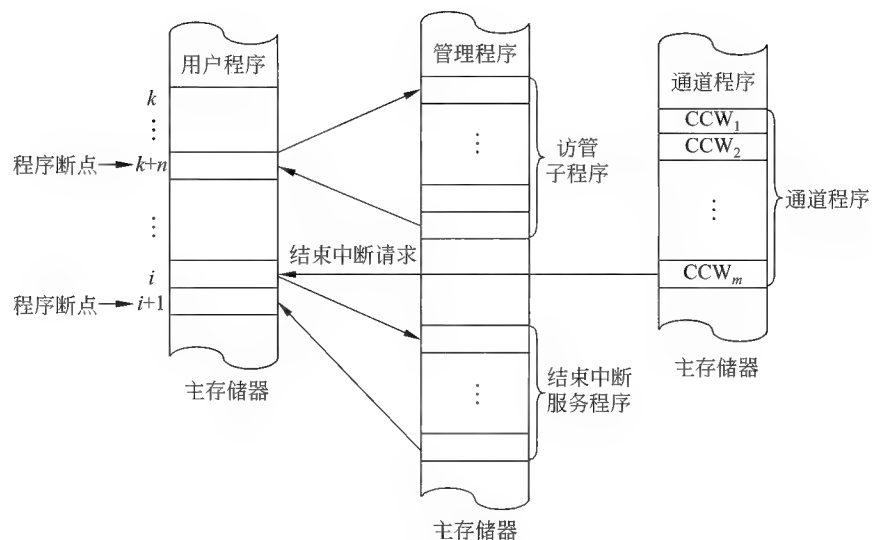


图 8.7 通道操作全过程

8.3 相关知识介绍

1. 输入输出系统的特点

在计算机系统中,通常把处理机与主存储器之外的部分统称为输入输出系统,输入输出系统是计算机系统中最具多样性和复杂性的部分。

输入输出系统的特点集中反映在异步性、实时性和与设备无关性这 3 项基本要求上,它们对输入输出系统的组织产生决定性的影响。

(1) 异步性

输入输出设备的工作在很大程度上独立于处理机之外,通常不使用统一的中央时钟,各个设备按照自己的时钟工作,但又要在某些时刻接受处理机的控制。

(2) 实时性

对于一般外部设备,处理机必须按照不同设备所要求的传送方式和传输速率不失时机地为设备提供服务,包括从设备接收数据,向设备发送数据及对设备的控制等。如果错过了服务的时机,就可能丢失数据或造成外设工作的错误。

用于实时控制的计算机系统对时间性的要求更强,如果处理机提供的服务不及时,很可能造成巨大的损失,甚至造成人身伤害。

对于计算机系统本身的硬件或软件错误:如电源故障、数据校验错、页面失效、非法指令、地址越界等,CPU 也必须及时处理。

(3) 与设备无关性

计算机系统为了能够适应各种外设的不同要求,规定了一些独立于具体设备的标准接口。例如,串行接口、并行接口、SCSI 接口、USB 接口等。各种外设必须根据自己的特点和要求,选择其中的一种标准接口与计算机连接。凡是连接到同一种标准接口上的不同类型的设备,它们之间的差异必须由设备本身的控制器通过硬件和软件来进行填补。这样,处理机本身就无需了解各种外设特定的具体工作细节,可以采用统一的硬件和软件对品种繁多的设备进行管理。

2. 主机与外设的连接方式

主机和外设的连接方式大致可以分为 3 类,辐射型连接、总线型连接和结合型连接,如图 8-8 所示。

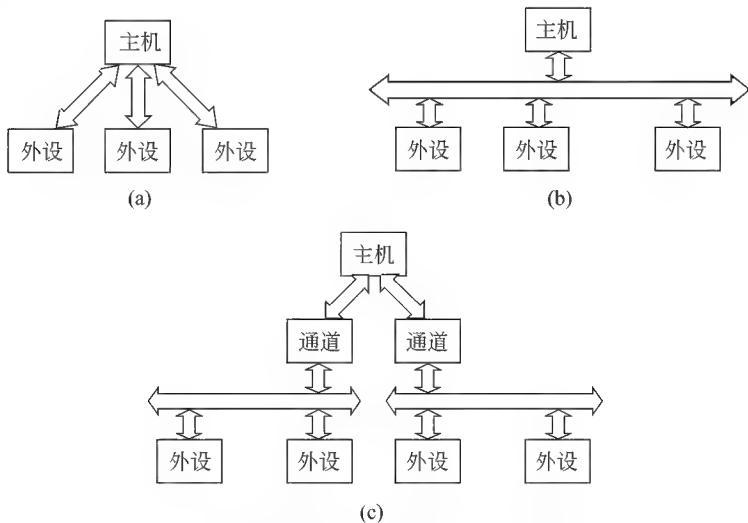


图 8-8 主机与外设间的连接方式

(1) 辐射型连接(星型)

图 8-8(a)中,主机和每个外设间都有各自独立的数据通路,形成以主机为中心向各个设备辐射的星形连接。这种连接方式具有控制简单的优点,但结构复杂、连线多,缺乏灵活性,尤其当外设数量较多时,连接起来可能很麻烦,因而现在已被淘汰。

(2) 总线型连接

图 8-8(b)中,主机通过系统总线与外设连接,各外设经三态门挂载在总线上,故称为总线型连接。这种连接方式具有结构简单,易于扩展等优点,而且各外设之间也有可能通过同一组总线直接通信。其缺点是所有的外设都通过同一组总线分时地工作,由于信息吞吐量有限,将影响交换速度。这种结构广泛用于微型机与小型机中。

(3) 结合型连接

图 8-8(c)所示的连接方式可看成前述两种方式的结合型。主机通过“通道”来管理外设的输入输出操作,主机与通道间采用辐射型连接,而通道和外设间则采用总线型连接。

3. 中断系统的软硬件功能分配

中断的全过程可以大致分为中断请求、中断判优、中断响应、中断处理和中断返回 5 个阶段,这 5 个阶段又可以细分出许多的功能。其中,有的功能必须用硬件来实现,有的功能必须由软件来实现,而大部分功能既可以用硬件实现,也可以用软件实现。因此,在设计一台计算机的中断系统时,如何适当地分配中断系统的软件和硬件功能,是设计好中断系统最关键的问题。中断系统中软件与硬件的功能分配主要考虑两个因素。

① 中断响应时间:从一个中断源发出中断请求到处理机响应这个中断源的中断请求,并开始执行这个中断源的中断服务程序所用的这一段时间称为中断响应时间。在中断系统中,中断响应时间是一个非常重要的指标。特别是在实时计算机系统中,中断响应时间是整个计算机系统的一个关键性指标。

② 灵活性:一般情况下,用硬件实现速度快,但灵活性差;用软件实现灵活性好,但速度低。

上述两个要求实际上是互相矛盾的。如果要减少中断响应时间,那么中断处理过程中那些既能用硬件实现,也能用软件实现的功能,要尽量用硬件来实现,但是这样做就必然失去了灵活性。相反,如果用软件实现的功能多了,灵活性虽然好了,但中断响应时间就必然增加。

细化之后的中断全过程如图 8-9 所示,不同计算机系统的中断全过程中的各种功能完成的顺序可能有所不同。例如,“转向中断服务程序入口”这一功能可以插在从“保存断点”之后到第一次“开中断”(保护现场后)之间的任何一个地方。

在图 8-9 所示的全部功能中,只有“保护断点”和“转向中断服务程序入口”这两个功能必须用硬件来实现(通过中断隐指令)。这是因为中断响应发生在现行程序的什么地方是不确定的,一般不能由程序员来安排。另外,第一次“关中断”一般也用硬件来实现。同样,也只有“中断服务”和“中断返回”这两个功能必须用软件来实现。其中“中断返回”需要执行一条中断返回指令。至于“中断服务”当然是要用软件来实现的,否则也就不能称为程序中断方式了。

中断源发出中断请求,
CPU 响应此中断请求

- 保存断点
- ◎ 关中断
- ◎ 撤销本次中断请求
- ◎ 识别中断源
- ◎ 改变设备的屏蔽状态
- 转向中断服务程序入口
- ◎ 保存现场
- 开中断
- 中断服务
- 关中断
- ◎ 恢复现场
- ◎ 恢复屏蔽状态
- ◎ 开中断
- 中断返回

注: ● 表示本行的功能一般用硬件来实现
○ 表示本行的功能一般用软件来实现
◎ 表示本行的功能可以用硬件来实现,也可以用软件来实现

图 8-9 细化之后的中断全过程

通常,希望中断响应时间尽可能短,如果中断响应时间过长,在实时控制系统中,很可能失去控制的时机或丢失控制信号;在数据采集或数据传输系统中,有可能丢失数据。影响中断响应时间的因素主要有4个:

① 最长指令执行时间。在一条指令执行期间,不允许被中断。由于中断源的中断请求是随机发出的,可能发生在一条指令执行过程中的任何时刻,考虑最坏情况,就是最长指令执行时间。

② 在一条指令执行完成后,处理其他更紧迫的任务所用的时间。例如,处理DMA请求等。

③ 从第一次“关中断”到第一次“开中断”所经历的时间。在整个中断响应时间中,这段时间往往是最主要的。如果这些要做的事情都用硬件来完成,中断响应时间就可以缩短很多。相反,如果其中的大部分功能都用软件来实现,则中断响应时间就会很长。

④ 多个中断源同时请求中断时,通过软件找到相关中断源的中断服务程序入口地址所经历的时间。

上述4部分时间中,第③部分是中断系统设计中需要考虑的主要问题。

4. 可屏蔽中断和不可屏蔽中断

在主教材中已经讨论了多种中断类型,除此以外,在许多系统中还可将中断分为可屏蔽中断和不可屏蔽中断两部分,例如在80x86系统中就有可屏蔽中断INTR和不可屏蔽中断NMI两种。

可屏蔽中断是指可不响应或暂不响应,或有条件响应的中断。当中断源产生中断时,用程序方法可以有选择的封锁部分中断,使之不发出中断请求,而允许其余部分中断发出中断请求。具体实现方法是,在硬件上为每个可屏蔽中断源设一个屏蔽触发器,用程序方法将该触发器置“1”,则相应中断源不能发出中断请求;若将其置“0”,则允许该中断源发出中断请求。

不可屏蔽中断是指必须立即响应、不能回避和禁止的中断。不可屏蔽的中断源产生的中断必须立即响应,即它们具有高的优先级,如断电中断是具有最高优先级的不可屏蔽中断,对断电中断的处理安排在DMA和所有中断之前。自愿中断也属于不可屏蔽中断。

5. 向量中断与向量地址的产生

为了提高CPU响应中断的速度,往往采用硬件排队的向量中断响应方法。在向量中断中,每个中断源都给出一个中断向量和向量地址。当CPU响应中断后,由中断机构自动地将向量地址通知CPU,由向量地址指明向量的位置并实现向量的切换,不必通过软件来查询中断源,这种响应称为向量中断响应。由于CPU每次只能为一个中断请求服务,因此同样也存在着优先级排队问题。此时,将各级设备向量地址形成电路和优先级排队电路集合在一起,组成向量中断优先权编码器。下面以8级中断为例说明硬件排队

向量中断的基本概念。假定 8 级中断请求为 $I_0 \sim I_7$, 并规定 0 级优先级最高, 1 级次之, 7 级最低。表 8-2 为 8 级中断屏蔽码和向量地址, 其中屏蔽码中 0 表示允许中断, 1 表示屏蔽中断。

表 8-2 8 级中断屏蔽码和向量地址

中断请求	屏蔽码								向量地址 VA		
	0	1	2	3	4	5	6	7	VA ₀	VA ₁	VA ₂
I ₀	1	1	1	1	1	1	1	1	0	0	0
I ₁	0	1	1	1	1	1	1	1	0	0	1
I ₂	0	0	1	1	1	1	1	1	0	1	0
I ₃	0	0	0	1	1	1	1	1	0	1	1
I ₄	0	0	0	0	1	1	1	1	1	0	0
I ₅	0	0	0	0	0	1	1	1	1	0	1
I ₆	0	0	0	0	0	0	1	1	1	1	0
I ₇	0	0	0	0	0	0	0	1	1	1	1

图 8-10 为向量地址形成电路。第一列为屏蔽寄存器 MASK 的各位, 第二列为中断请求寄存器 INTR 的各位, $D_0 \sim D_7$ 为各设备的完成信号; 第三列“与门”的输出是各级中断请求信号的输出, 并将其加入到并行排队器; 第四列是并行优先权排队电路; 第五列由三个“或门”组成 8-3 编码器, 输出为向量地址 VA_0 、 VA_1 、 VA_2 。图 8-10 中虚线框内称为向量中断优先权编码器。所有中断向量的集合组成一张向量表, 向量表建立在主存中。向量地址可用主存地址 MAR 表示, 这时所形成的 VA_0 、 VA_1 、 VA_2 三位变量嵌入在 MAR 中。如果向量表是连续表, 而且存放在主存从 0 号单元开始的区域, 则 VA_0 、 VA_1 、 VA_2 就作为 MAR 的低位 MAR_2 、 MAR_1 、 MAR_0 ; 若向量表在主存中是分散分布, 则向量地址嵌入主存地址的中间部分。

例 8-1 若处理机给出的屏蔽码为 11000011, 在某一个时刻 t , D_1 、 D_3 、 D_4 这 3 个设备的完成信号同时到达, 此时向量地址应该是什么?

解: 因为各设备的向量地址又是中断源的优先级, D_1 、 D_3 、 D_4 这 3 个设备的优先级应该是 $1 > 3 > 4$ 。屏蔽码为 11000011, 它屏蔽了 0 号、1 号、6 号、7 号设备的中断请求, 但对 2 号、3 号、4 号、5 号设备是开放的。对于 1 号、3 号、4 号设备来说, 1 号设备中断请求应被屏蔽, 3 号、4 号设备既满足了外设工作已完成的条件, 而且设备未被屏蔽, 所以可有中断请求参与排队, 又因 3 号设备优先级大于 4 号设备, 所以经排队线路应产生 3 号设备的向量地址 011。

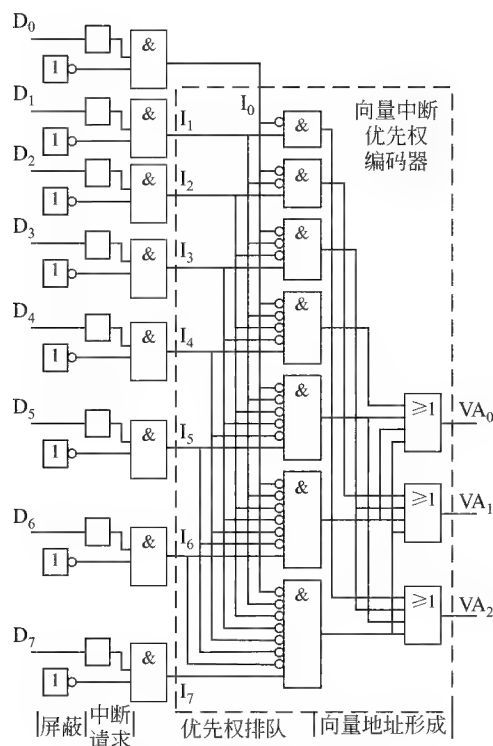


图 8 10 向量地址形成电路

6. 向量中断的执行过程

图 8-11 为向量中断的执行过程。它表示优先级为 6 级的中断源打断优先级为 7 级的程序的执行过程。

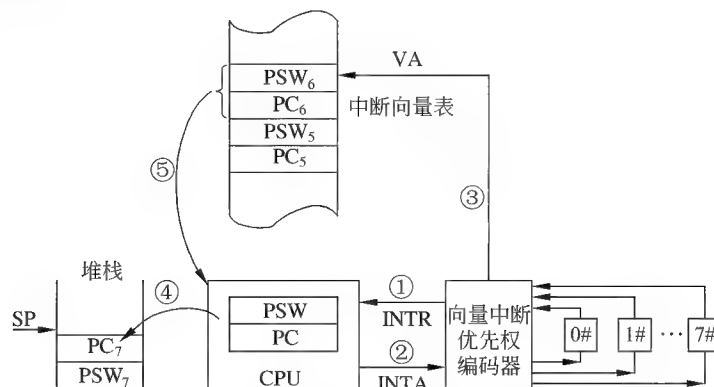


图 8 11 向量中断的执行过程

向量中断过程如下：

- ① 当设备有中断源请求时,通过中断请求线 INTR 向 CPU 提出申请;
- ② CPU 响应后,由中断响应线 INTA 回答向量中断优先权编码器;
- ③ 由向量中断优先权编码器形成优先级最高的向量地址 VA,通知 CPU;
- ④ CPU 执行中断隐指令,将 PSW₇ 和 PC₇ 压入堆栈,并用 PSW₆ 和 PC₆ 填写 CPU 内的 PSW 和 PC,致使 CPU 脱离现行程序,转去执行中断服务程序。

中断隐指令执行后,PC 中存放着优先权最高的外设中断服务程序的入口地址,从而使 CPU 运行中断服务程序。

返回时,中断服务程序的末尾是一条中断返回指令(RETI)。RETI 指令从堆栈中取出断点 PC₇ 和程序状态字 PSW₇,即可以返回原来的程序继续执行。

7. 中断升级的另一种方法 —— 改变处理机优先级

在主教材中已经讨论了中断升级的方法 —— 每个中断源设置一个中断屏蔽位,通过改变中断屏蔽码,可以改变中断源的中断处理次序。下面介绍中断升级的另一种方法 —— 改变处理机优先级。

中断优先级不仅在处理机响应中断源的中断请求时使用,而且为每个中断源的中断服务程序也赋予同样的中断优先级。如果一台处理机共有 n 个中断源,则在处理机的状态字中需要设置 $\log_2(n+1)$ 个中断屏蔽位。这里,不把它称为中断屏蔽码而称为中断优先级。处理机本身的优先级一般设置为最低,通常,处理机在运行现行程序时,其优先级为 0 级。另外,要为每一个中断源分别建立处理机状态字,通常把它们存放在主存的一个固定区域中。这些中断源的处理机状态字中同样也有一个中断优先级字段,而且每个中断优先级一般都可以由程序员通过软件来进行修改。

处理机在响应某一个中断源的中断请求后,就把属于这个中断源的处理机状态字作为当前处理机的状态字,这时候处理机的优先级也就改变成为程序员为这个中断源设置的中断优先级。这时,只有中断优先级高于当前处理机优先级的中断源才能中断当前的中断服务程序。

正常工作的情况下,在各个中断源的处理机状态字中设置的中断优先级应该与这个中断源本身的硬件优先级相同。这时,处理机响应中断源的中断请求和完成中断服务的过程将严格按照中断源的硬件优先级进行。

如果要改变中断源的中断处理次序,即在有多个中断源同时请求中断时,让某些硬件中断优先级较低的中断源先得到处理机的服务,可以通过修改相关中断源的处理机状态字来实现。

例 8-2 某处理机共有 4 个中断源 D₁、D₂、D₃ 和 D₄,它们在串行排队链中的硬件中断优先级从低到高分别为 1 级、2 级、3 级和 4 级。处理机本身的优先级最低,为 0 级。在中断源 D₁、D₂、D₃、D₄ 的处理机状态字中,程序员为它们设置的优先级分别为 4 级、3 级、2

级、1级。

解：因为有4个中断源，因此在处理机状态字中要设置3个中断屏蔽位。其中，000为处理机本身的优先级，001~100分别表示4个中断源的中断优先级。如果当处理机正在执行现行程序时，4个中断源同时请求中断服务，图8-12给出处理机实际响应中断源的中断请求和完成中断服务的过程。

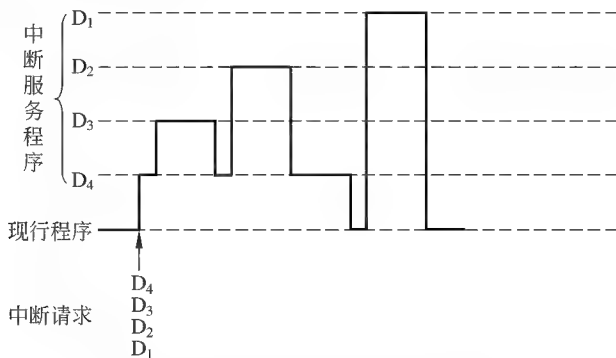


图 8-12 采用处理机优先级进行中断升级

当处理机运行现行程序时，处理机的优先级最低，为0级。4个中断源同时请求中断，通过硬件排队器选择其中硬件中断优先级最高的中断源 D_4 ，处理机实现响应它的中断请求，并且处理机的优先级改变为1级。这时，由于 D_1 、 D_2 、 D_3 这3个中断源仍然在请求中断，而且它们的硬件中断优先级都比当前处理机的优先级高，因此，处理机要响应其中硬件优先级最高的中断源 D_3 的中断请求，并且把处理机的优先级改变为2级。这时，还剩下 D_1 和 D_2 两个中断源在请求中断，它们的硬件中断优先级都不高于当前处理机的优先级，因此，中断源 D_3 的中断服务程序能够一直执行完成。当处理机从中断源 D_3 的中断服务程序返回到中断源 D_4 的中断服务程序中时，处理机的中断优先级又变成1级。这时，还没有得到处理机响应的两个中断源 D_1 和 D_2 中，只有中断源 D_2 的硬件优先级高于当前处理机的中断优先级。因此，处理机将立即响应中断源 D_2 的中断请求，又把处理机的优先级改变为3级。由于剩下的中断源 D_1 的硬件优先级低于当前处理机的优先级，处理机将把中断源 D_2 的中断服务程序全部执行完毕才会返回到中断源 D_4 的中断服务程序中。这时，最后一个没有被处理机响应的中断源 D_1 的硬件中断优先级与当前处理机的优先级相同，要等待处理机把中断源 D_4 的中断服务程序全部执行完成，并返回到现行程序之后，中断源 D_1 的中断请求才能得到处理机的响应。由于中断源 D_1 是最后一个被处理机响应并得到服务的中断源，因此，它的中断服务程序能够一直执行完成。在中断源 D_1 的中断服务程序全部执行完成，处理机返回执行现行程序时，全部中断源的中断请求也就处理完成了。

从图 8-12 中看到,虽然 4 个中断源 D_1 、 D_2 、 D_3 、 D_4 的硬件中断优先级是从低到高顺序排列的,并且 4 个中断源同时请求中断,但是,通过改变各个中断源所属的处理机状态字内的中断优先级,使得处理机实际完成中断处理的顺序改变成了 D_3 、 D_2 、 D_4 、 D_1 。当然,还可以根据需要,任意改变中断处理的次序。

改变处理机优先级的方法与每一个中断源都设置一个屏蔽位的方法相比,两者都能由程序员通过软件来改变中断源的中断处理次序,因此,它们都具有灵活性好的特点。它们的差别有两个,第一,两者使用的概念不同,前者使用的是中断屏蔽码,后者使用的是中断优先级;第二,前者要为每一个中断源设置一个中断屏蔽位,所需要的位数比较多,而后者表示优先级所使用的位数要少得多。当然,前者使用起来要比后者方便些,例如,前者可以任意屏蔽掉一个或几个中断源,而后者只能屏蔽掉比某一个中断优先级低的中断源。

8. 连接多台外设的 DMA 控制器

主教材中讨论的 DMA 控制器只能连接一台外设,故又称为单通道 DMA 控制器。如果一个 DMA 控制器连接着多台设备,则 DMA 控制器需要有些变化,可能采用选择型也可能采用多路型。

(1) 选择型 DMA 控制器

选择型 DMA 控制器虽然在物理上可连接多台高速外设,但在逻辑上只允许连接一台外设。各外设通过一个简单的 I/O 总线与 DMA 控制器某段时间内,只有被选中的一台设备使用局部 I/O 总线。

选择型 DMA 控制器适合于数据传输率很高,以致接近于主存速度的设备,在这种情况下,不允许在批量传送中切换设备。选择型 DMA 控制器的功能相当于一个数据块传送的切换开关,以数据块为单位进行选择与切换。

(2) 多路型 DMA 控制器

多路型 DMA 控制器不仅在物理上可连接多台外设,而且在逻辑上也允许多台外设同时工作,各外设采用字节或字交叉方式进行数据传送。这种类型由一个 DMA 控制器和多个接口组成,有多少个接口,就可连接多少台设备。

多路型 DMA 控制器适合于连接多个数据传输率不很高的外设,实用的多路型 DMA 控制器往往兼有选择型 DMA 控制器的功能。如果采用单字传送方式,就是典型的多路型;如果采用成组连续传送,让各设备以数据块为单位,则是选择型。

9. 通道中的数据传送过程

一个字节多路通路是分时为多台低速和中速外设服务的,在有 P 台设备同时连接到一个字节多路通道上时,它们的数据传送过程如图 8-13(a)所示。

下面是在图 8-13(a)中每一个参数的含义。

T_s : 设备选择时间。从通道需要设备发出数据传送请求开始,到通道实际为这台设备传送数据所需要的时间。

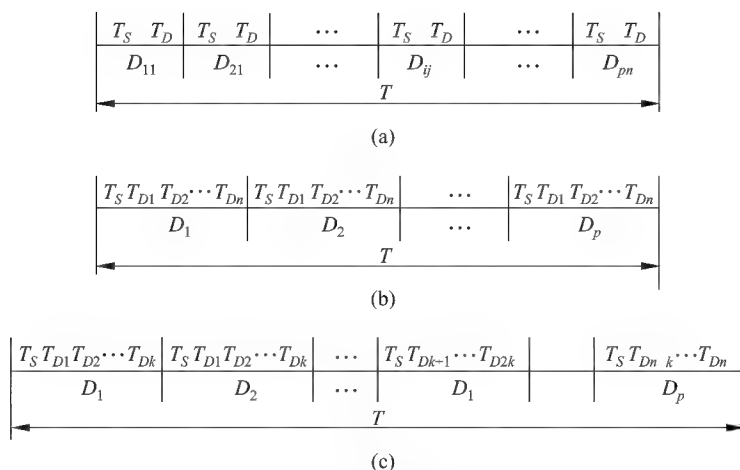


图 8 13 通道的数据传送过程

T_D : 传送一个字节所用的时间,实际上就是通道执行一条通道指令,即数据传送指令所用的时间。

P : 在一个通道上连接的设备台数,且这些设备同时都在工作。

n : 每一个设备传送的字节个数,这里,假设每一台设备传送的字节数都相同,都是 n 个字节。

D_{ij} : 连接在通道上的第 i 台设备传送的第 j 个数据,其中 $i=1,2,\dots,p,j=1,2,\dots,n$ 。

T : 通道完成全部数据传送工作所需要的时间。

在字节多路通道中,通道每连接一台外设,只传送一个字节,然后又与另一台设备相连接,因此, T_s 和 T_D 是间隔进行的。

当一个字节多路通道上连接有 P 台设备,每台设备都传送 n 个字节时,所需要的时间为:

$$T_{\text{byte}} = (T_s + T_D) \cdot P \cdot n$$

选择通道在一段时间内只能单独为一台高速外设服务,当这台设备的数据传送工作全部完成后,通道才能为另一台设备服务。选择通道的数据传送过程如图 8-13(b)所示,图中除去与字节多路通路相同的参数以外,还有下面参数。

T_{Di} : 通道传送第 i 个数据所用的时间,其中 $i=1,2,\dots,n$ 。

D_i : 通道正在为第 i 台设备服务, $i=1,2,\dots,p$ 。

在选择通道中,通道每连接一台外设,就把这个设备的 n 个字节全部传送完成,然后再与另一台设备相连接,因此,在一个 T_s 之后,有连续 n 个数据传送时间 T_D 。

当一个选择通道连接 P 台设备,每台设备都传送 n 个字节时,所需要的时间为:

$$T_{\text{select}} = \left| \frac{T_s}{n} + T_D \right| \cdot P \cdot n$$

数组多路通道在一段时间内只能为一台高速设备传送数据,但同时可以有多台高速设备在寻址,包括定位和找扇区。数组多路通道的数据传送过程如图 8-13(c)所示,图中所用参数与前两种类型相同,另外还有下面参数。

k :一个数据块中的字节个数。在一般情况下, $k < n$ 。

数据多路通道每连接一台高速设备,一般传送一个数据块,传送完成后,又与另一台高速设备连接,再传送一个数据块,因此,在一个 T_s 之后,有连续 k 个数据传送时间 T_D 。

当一台数组多路通道连接 P 台设备,每台设备都传送 n 个字节时,所需要的时间为:

$$T_{\text{block}} = \left| \frac{T_s}{k} + T_D \right| \cdot P \cdot n$$

10. 通道的流量分析

通道流量是指通道在数据传送期内,单位时间里传送的字节数。它能达到的最大流量称为通道极限流量。

假设通道选择一次设备的时间为 T_s ,每传送一个字节的时间为 T_D ,通道工作时的极限流量分别为:

字节多路通道

$$f_{\text{max}\cdot\text{byte}} = \frac{P \cdot n}{(T_s + T_D) \cdot P \cdot n} = \frac{1}{T_s + T_D}$$

每选择一台设备只传送一个字节。

选择通道

$$f_{\text{max}\cdot\text{select}} = \frac{P \cdot n}{\left| \frac{T_s}{n} + T_D \right| \cdot P \cdot n} = \frac{1}{\frac{T_s}{n} + T_D} = \frac{n}{T_s + nT_D}$$

每选择一台设备就把 n 个字节全部传送完。

数组多路通道

$$f_{\text{max}\cdot\text{block}} = \frac{P \cdot n}{\left| \frac{T_s}{k} + T_D \right| \cdot P \cdot n} = \frac{1}{\frac{T_s}{k} + T_D} = \frac{k}{T_s + kT_D}$$

每选择一台设备传送定长 k 个字节。

若通道上接 P 台设备,则通道要求的实际流量分别为:

字节多路通道

$$f_{\text{byte}} = \sum_{i=1}^P f_i$$

即所接 P 台设备的速率之和。

选择通道

$$f_{\text{select}} = \max_{i=1}^P f_i$$

数组多路通道

$$f_{\text{block}} = \max_{i=1}^P f_i$$

即所接 P 台设备中速率最高者。

为使通道所接外部设备在满负荷工作时仍不丢失信息,应使通道的实际最大流量不能超过通道的极限流量。

如果在 I/O 系统中有多通道,各个通道是并行工作的,则 I/O 系统的极限流量应当是各通道或各子通道工作时的极限流量之和。

例 8-3 一个字节多路通道连接 D_1 、 D_2 、 D_3 、 D_4 、 D_5 共 5 台设备,这些设备分别每 $10\mu\text{s}$ 、 $30\mu\text{s}$ 、 $30\mu\text{s}$ 、 $50\mu\text{s}$ 和 $75\mu\text{s}$ 向通道发出一次数据传送的服务请求,请回答下列问题:

(1) 计算这个字节多路通道的实际流量和工作周期。

(2) 如果设计字节多路通道的最大流量正好等于通道实际流量,并假设对数据传输率高的设备,通道响应它的数据传送请求的优先级也高。5 台设备在 0 时刻同时向通道发出第一次传送数据的请求,并在以后的时间里按照各自的数据传输率连续工作。画出通道分时为每台设备服务的时间关系图,并计算这个字节多路通道处理完各台设备的第一次数据传送请求的时刻。

(3) 从时间关系图上可以发现什么问题? 如何解决这个问题?

解: 这个字节多路通道的时间流量为:

$$f_{\text{byte}} = \left| \frac{1}{10} + \frac{1}{30} + \frac{1}{30} + \frac{1}{50} + \frac{1}{75} \right| \text{MB/s} = 0.2 \text{MB/s}$$

通道的工作周期为:

$$T = \frac{1}{f_{\text{byte}}} = 5\mu\text{s}$$

包括设备选择时间 T_s 和传送一个字节的的时间 T_D 。

5 台设备向通道请求传送和通道为它们服务的时间关系如图 8-14 所示,向上的箭头表示设备的数据传送请求,有阴影的长方形表示通道响应设备的请求并为设备服务所用的工作周期。

在图 8-14 中,5 台设备在 0 时刻同时向字节多路通道发出第一次传送时间的请求,通道处理完各设备第一次请求的时间分别为:

处理完设备 D_1 的第一次请求的时刻为 $5\mu\text{s}$;

处理完设备 D_2 的第一次请求的时刻为 $10\mu\text{s}$;

处理完设备 D_3 的第一次请求的时刻为 $20\mu\text{s}$;

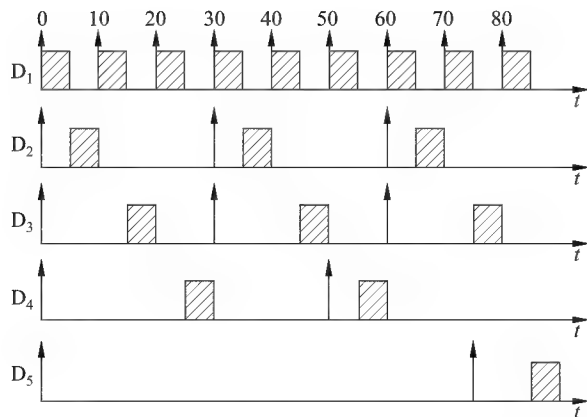


图 8-14 字节多路通道响应设备请求和为设备服务的时间关系图

处理完设备 D_4 的第一次请求的时刻为 $30\mu s$;

设备 D_5 的第一次请求没有得到通道的响应,直到第 $85\mu s$ 通道才开始响应设备 D_5 的服务请求,这时,设备已经发出了两个传送数据的服务请求,因此第一次传送的数据有可能丢失。

当字节多路通道的最大流量与连接在这个通道上的所有设备的数据流量之和非常接近时,虽然能够保证在宏观上通道不丢失设备的信息,但不能保证在某个局部时刻不丢失信息。由于高速设备在频繁地发出要求传送数据的请求时,总是被优先得到响应和处理,这就可能使低速设备的信息一时得不到处理而丢失,如本例中的设备 D_5 。为了保证本例中的字节多路通道能正常工作,可以采取以下措施来解决:

① 增加通道的最大流量,以保证连接在通道上的所有设备的数据传送请求能够及时得到通道的响应。

② 动态改变设备的优先级。例如,在图 8-14 中,只要在 $30\mu s \sim 70\mu s$ 之间临时提高设备 D_5 的优先级,就可使设备 D_5 的第一次传送请求及时得到通道的响应,其他设备的数据传送请求也能正常得到通道的响应。

③ 增加一定数量的数据缓冲器,特别是对优先级比较低的设备。例如,在图 8-14 中,只要为设备 D_5 增加一个数据缓冲器,它的第一次数据传送请求可在 $85\mu s$ 处得到通道的响应,第二次数据传送请求可以在 $145\mu s$ 处得到通道的响应,所有设备的数据都不会丢失。

8.4 教材习题解答

1. 什么是计算机的输入输出系统? 输入输出设备有哪些编址方式? 有什么特点?

解: 计算机的输入输出系统包括输入输出接口和输入输出信息传送控制方式等,它

们是整个计算机系统中最具有多样性和复杂性的部分。

输入输出设备有两种编址方式:I/O映射方式(独立编址)和存储器映射方式(统一编址),独立编址的优点是I/O指令和访存指令容易区分,外设地址线少,译码简单,主存空间不会减少,缺点是控制线增加了I/O读和I/O写信号;统一编址的优点是总线结构简单,全部访存类指令都可用于控制外设,可直接对外设寄存器进行各种运算,占用主存一部分地址,缩小了可用的主存空间。

2. 什么是I/O接口? I/O接口有哪些特点和功能? 接口有哪些类型?

解: I/O接口是主机和外设之间的交接界面,通过接口可以实现主机和外设之间的信息交换。接口的基本功能有:实现主机和外设的通信联络控制;进行地址译码和设备选择;实现数据缓冲;完成数据格式的变换;传递控制命令和状态信息。接口按数据传送方式分类,有串行接口和并行接口;按控制方式分类,有程序查询接口、程序中断接口、DMA接口;按灵活性分类,有可编程接口和不可编程接口;按通用性分类,有通用接口和专用接口;按I/O信号分类,有数字接口和模拟接口;按应用分类,有运行辅助接口、用户交互接口、传感接口、控制接口。

3. 并行接口和串行接口实质上的区别是什么? 其界面如何划分? 各有什么特点?

解: 有串行接口和并行接口。两者的实质区别在于外设和接口一侧的传送方式不同,而在主机和接口一侧,数据总是并行传送的。在并行接口中,外设和接口间的传送宽度是一个字节(或字)的所有位,一次传输的信息量大,但数据线的数目将随着传送数据宽度的增加而增加。在串行接口中,外设和接口间的数据是一位一位串行传送的,一次传输的信息量小,但只需一根数据线。在远程终端和计算机网络等设备离主机较远的场合下,用串行接口比较经济划算。

4. I/O数据传送可以采用哪些方式? 它们各有什么特点及应用场所? 试比较之。

解: 主机和外设之间的信息传送控制方式,经历了由低级到高级、由简单到复杂、由集中管理到各部件分散管理的发展过程,按其发展的先后次序和主机与外设并行工作的程度,可以分为4种。

程序查询方式是主机与外设间进行信息交换的最简单方式,输入和输出完全是通过CPU执行程序来完成的。这种方式控制简单,但外设和主机不能同时工作,系统效率很低,因此,仅适用于外设的数目不多、对I/O处理的实时要求不高、CPU的操作任务比较单一、并不很忙的情况。

程序中断方式无需等待查询,外设在做好输入输出准备时,向主机发中断请求,主机接到请求后就暂时中止原来执行的程序,转去执行中断服务程序对外部请求进行处理,在中断处理完毕后返回原来的程序继续执行。程序中断不仅适用于外部设备的输入输出操作,也适用于对外界发生的随机事件的处理。由于完成一次程序中断还需要许多辅助操作,因此主要适用于中、低速外设。

DMA 方式是在主存和外设之间开辟直接的数据通路,可以进行基本上不需要 CPU 介入的主存和外设之间的信息传送,这样不仅能保证 CPU 的高效率,而且能满足高速外设的需要。DMA 方式只能进行简单的数据传送操作,在数据块传送的起始和结束时还需 CPU 及中断系统进行预处理和后处理。

I/O 通道控制方式是 DMA 方式的进一步发展,在系统中设有通道控制部件,每个通道挂若干外设,主机在执行 I/O 操作时,只需启动有关通道,通道将执行通道程序,从而完成 I/O 操作。

5. 程序查询方式、程序中断方式、DMA 方式各自适用什么范围? 下面这些结论正确吗? 为什么?

(1) 程序中断方式能提高 CPU 利用率,所以在设置了中断方式后就没有再应用程序查询方式的必要了。

(2) DMA 方式能处理高速外部设备与主存间的数据传送,高速工作性能往往能覆盖低速工作要求,所以 DMA 方式可以完全取代程序中断方式。

解: 程序查询方式、程序中断方式、DMA 方式各自适用的范围见前述。

(1) 不正确。程序查询方式接口简单,可用于外设与主机速度相差不大,且外设数量很少的情况。

(2) 不正确。DMA 方式用于高速外部设备与主存间的数据传送,但 DMA 结束时仍需程序中断方式做后处理。

6. 什么是程序查询 I/O 传送方式? 试举例说明其工作原理,它有哪些优缺点?

解: 程序查询方式是主机与外设间进行信息交换的最简单方式,程序查询方式的核心问题在于需要不断地查询 I/O 设备是否准备就绪。

CPU 利用程序查询方式从硬盘上读取一个数据的过程是:CPU 首先启动键盘工作,然后测试键盘状态,若键盘数据未准备就绪,则输入缓冲寄存器的内容不可以使用,继续查询;若键盘数据已准备就绪,则执行输入指令取走该数据。这种方式的优点是控制简单,节省硬件,缺点是系统效率低。

7. 图 8-5(主教材第 293 页)是以程序查询方式实现与多台设备进行数据交换的流程图,试分析这种处理方式存在的问题以及改进措施。

解: 若有多个外设需要用查询方式工作时,CPU 巡回检测各个外设,逐个进行查询,发现哪个外设准备就绪,就对该外设实施数据传送,然后再对下一个外设查询,依次循环。在整个查询过程中,CPU 不能做别的事。如果某一外设刚好在查询过自己之后才处于就绪状态,那么它就必须等 CPU 查询完其他外设后再次查询自己时,才能让 CPU 为它服务,这对于实时性要求较高的外设来说,就可能丢失数据。改进的措施可以采用增加缓冲寄存器的方法。

8. 如果采用程序查询方式从磁盘上输入一组数据,设主机执行指令的平均速度为

100 万条指令每秒,试问从磁盘上读出相邻两个数据的最短允许时间间隔是多少?若改为中断式输入,这个间隔是更短些还是更长些?由此可得出什么结论?

解:指令的平均执行时间为 $1\mu\text{s}$,若采用程序查询方式,每传送一个数据至少需要 5 条指令,则从磁盘上读出相邻两个数据的最短允许时间间隔为 $5\mu\text{s}$ 。若改为中断方式输入,这个间隔不会缩短,只会延长。由此可知,中断方式并不适合磁盘这类高速外设使用。

9. 什么是中断? 外部设备如何才能产生中断?

解:中断是指:计算机执行现行程序的过程中,出现某些急需处理的异常情况和特殊请求,CPU 暂时中止现行程序,而转去对随机发生的更紧迫的事件进行处理,在处理完毕后,CPU 将自动返回原来的程序继续执行。

外部设备(中断源)准备就绪后会主动向 CPU 发出中断请求。通常由外设的完成信号将相应的中断请求触发器置成“1”状态,表示该中断源向 CPU 提出中断请求。

10. 中断为什么要判优? 有哪些具体的判优方法? 各有什么优缺点?

解:当多个中断源同时发出中断请求时,CPU 在任何瞬间只能响应一个中断源的请求,所以要把全部中断源按中断的性质和轻重缓急安排优先级,以保证响应优先级别最高的中断请求。中断判优的方法可分为:软件判优法和硬件判优法。前者简单,可以灵活地修改中断源的优先级别,但查询、判优完全是靠程序实现的,不但占用 CPU 时间,而且判优速度慢。后者可节省 CPU 时间,速度快,但是需要硬件判优电路,成本较高。

11. CPU 响应中断应具备哪些条件?

解:① CPU 接收到中断请求信号;

② CPU 允许中断;

③ 一条指令执行完毕。

12. 什么叫中断隐指令? 中断隐指令有哪些功能? 中断隐指令如何实现?

解:CPU 响应中断之后,经过某些操作,转去执行中断服务程序。这些操作是由硬件直接实现的,称为中断隐指令。中断隐指令并不是指令系统中的一条真正的指令,它没有操作码,所以中断隐指令是一种不允许、也不可能为用户使用的特殊指令。其所完成的操作主要有:

① 保存断点;

② 暂不允许中断;

③ 引出中断服务程序。

13. 什么是中断向量? 中断向量如何形成? 向量中断和非向量中断有何差异?

解:中断向量是指向量中断在中断事件在提出中断请求时,通过硬件向主机提供的中断向量地址。中断向量由中断源的有关硬件电路形成。

向量中断和非向量中断的区别在于:前者是指那些中断服务程序的入口地址是由中断事件自己提供的中断;后者是指中断事件不能直接提供中断服务程序入口地址的中断。

14. 在程序中断处理中,要做到现行程序向中断服务程序过渡和中断服务程序执行完毕返回现行程序,必须进行哪些关键性操作?一般采用什么方法实现这些操作?

解: 最关键的操作有保存断点,适时开、关中断,保护和恢复现场等,其中部分工作由硬件完成,部分工作由软件完成。

15. 假定某机的中断处理方式是将断点存入 00000Q 单元,并从 77777Q 单元取出指令(即中断服务程序的第一条指令)执行。试排出完成此功能的中断周期微操作序列,并判断出中断服务程序的第一条指令是何指令(假定主存容量为 2^{15} 个单元)?

解: 中断周期微操作序列:

00000Q → MAR

(PC) → MDR

WRITE

0 → EINT

77777Q → PC

中断服务程序的第一条指令必须是一条无条件转移指令,否则 PC+1 将会变为 00000Q,断点被当成指令。

16. 假设有 1、2 两个设备,其优先级为设备 1 > 设备 2,若它们同时提出中断请求,试说明中断处理过程,画出其中断处理过程示意图,并标出断点。

解: 中断处理过程示意图见图 8-15。

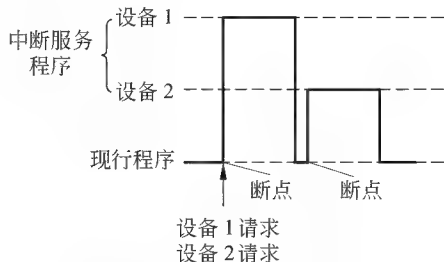


图 8 15 中断处理过程示意图

17. 现有 A、B、C、D 共 4 个中断源,其优先级由高向低按 A、B、C、D 顺序排列。若中断服务程序的执行时间为 $20\mu s$,请根据图 8-16 所示时间轴给出的中断源请求中断的时刻,画出 CPU 执行程序的轨迹。

解: CPU 执行程序的轨迹如图 8-17 所示。

18. 设某机有 5 级中断: L_0 、 L_1 、 L_2 、 L_3 、 L_4 ,其中断响应优先次序为: L_0 最高、 L_1 次之、……、 L_4 最低。现在要求将中断处理次序改为 $L_1 \rightarrow L_3 \rightarrow L_0 \rightarrow L_4 \rightarrow L_2$,试问:

(1) 各级中断服务程序中的各中断屏蔽码应如何设置(设每级对应一位,当该位为“0”,表示中断允许;当该位为“1”,表示中断屏蔽)?

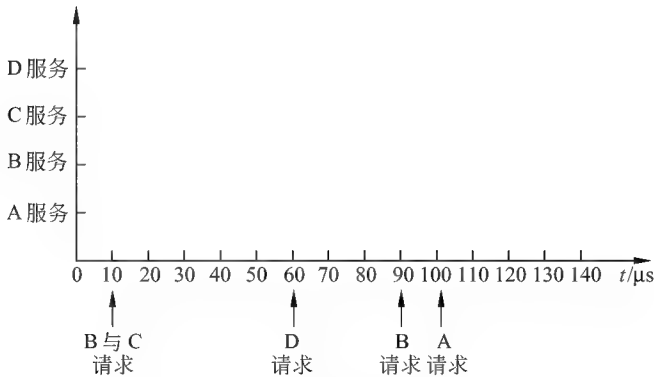


图 8 16 中断请求时间轴

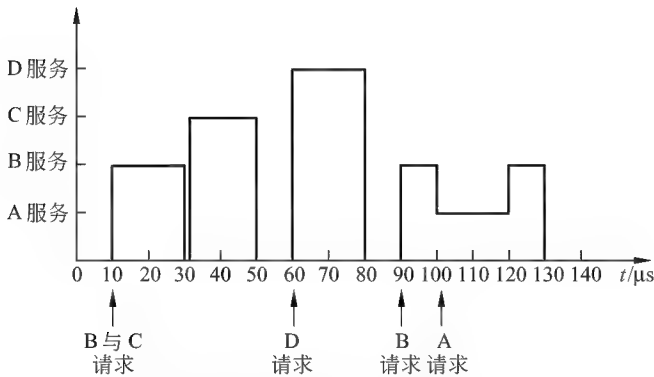


图 8 17 CPU 执行程序的轨迹

(2) 若这 5 级同时都发出中断请求,试画出进入各级中断处理过程示意图。

解：(1) 各级中断服务程序中的各中断屏蔽码设置如表 8-3 所示。

表 8 3 中断屏蔽码

程序级别	屏 蔽 码				
	0 级	1 级	2 级	3 级	4 级
第 0 级	1	0	1	0	1
第 1 级	1	1	1	1	1
第 2 级	0	0	1	0	0
第 3 级	1	0	1	1	1
第 4 级	0	0	1	0	1

(2) 5 级中断同时发出中断请求, 各级中断处理过程示意如图 8-18 所示。

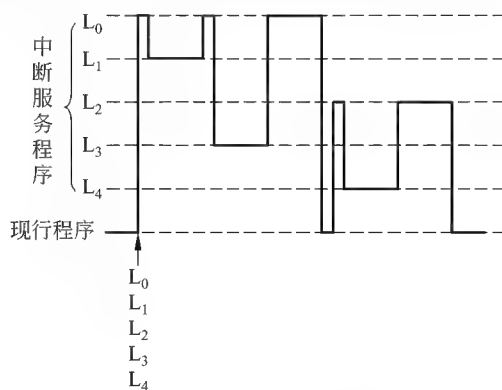


图 8 18 中断处理过程示意

19. 实现多重中断应具备何种条件? 如有 A、B、C、D 共 4 级中断, A 的优先级最高, B 次之, …… , D 最低。如在程序执行过程中, C 和 D 同时申请中断, 该先响应哪级中断? 如正在处理该中断时, A、B 又同时有中断请求, 试画出该多级中断处理的流程来。

解: 多重中断又称为中断嵌套, 在执行某个中断服务程序的过程中, CPU 可去响应级别更高的中断请求。

多级中断处理的流程如图 8-19 所示。

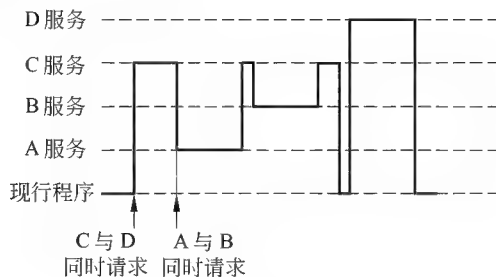


图 8 19 多级中断处理的流程

20. CPU 响应 DMA 请求和响应中断请求有什么区别? 为什么通常使 DMA 请求的优先级高于中断请求?

解: 对中断请求的响应时间只能发生在每条指令执行完毕时, 而对 DMA 请求的响应时间可以发生在每个机器周期结束时。DMA 方式常用于高速外设的成组数据传送, 如果不及时处理将丢失信息, 所以 DMA 请求的优先级应高于中断请求。

21. 什么是 DMA 传送方式? 试比较常用的 3 种 DMA 传送方法的优缺点?

解: DMA 传送方式是在外设和主存之间开辟一条“直接数据通道”, 在不需 CPU

干预也不需要软件介入的情况下在两者之间进行的高速数据传送方式。

常用的 DMA 传送方法有:CPU 停止访问主存法、存储器分时法和周期挪用法。

CPU 停止访问主存法适用于高速外设的成组传送,可以减少系统总线控制权的交换次数,有利于提高输入输出的速度。

存储器分时法无需申请和归还总线,可在 CPU 不知不觉中进行 DMA 传送;但这种方法需要主存在原来的存取周期内为两个部件服务,如果要维持 CPU 的访存速度不变,就要求主存的工作速度提高一倍。另外,由于大多数外设的速度都不能与 CPU 相匹配,所以供 DMA 使用的时间片可能成为空操作,将会造成一些不必要的浪费。

周期挪用法是前两种方法的折中。

22. 实现 DMA 传送需要哪些硬件支持?

解: DMA 传送需要 DMA 控制器的支持。DMA 控制器中主要有下面各部分。

主存地址计数器:用来存放待交换数据的主存地址。

传送长度计数器:用来记录传送数据块的长度。

数据缓冲寄存器:用来暂存每次传送的数据。

DMA 请求触发器:每当外设准备好数据后给出一个控制信号,使 DMA 请求触发器置位。

控制/状态逻辑:用于指定传送方向,修改传送参数,并对 DMA 请求信号和 CPU 响应信号进行协调和同步。

中断机构:当一个数据块传送完毕后触发中断机构,向 CPU 提出中断请求,CPU 将进行 DMA 传送的结束处理。

23. 简述 DMA 传送的工作过程。

解: DMA 传送的工作过程为:

① DMA 预处理:在 DMA 传送之前必须要做准备工作,即初始化。CPU 首先执行几条 I/O 指令,用于测试外设的状态、向 DMA 控制器的有关寄存器设置初值、设置传送方向、启动该外部设备等。在这些工作完成之后,CPU 继续执行原来的程序,由外设向 DMA 控制器发 DMA 请求,再由 DMA 控制器向 CPU 发总线请求。

② 数据传送:DMA 的数据传送可以单字节(或字)为基本单位,也可以数据块为基本单位。

③ DMA 后处理:当传送长度计数器计到 0 时,DMA 操作结束,DMA 控制器向 CPU 发中断请求,CPU 停止原来程序的执行,转去执行中断服务程序做 DMA 结束处理工作。

24. 在主存接收从磁盘送来的一批信息时:

(1) 假定主存的周期为 $1\mu\text{s}$,若采用程序查询方式传送,试估算在磁盘上相邻两数据字间必须具有的最短允许时间间隔是多少?

(2) 若改为中断方式传送,这个时间又会怎样? 是否还有更好的传送方式?

(3) 在采用更好的传送方式下,假设磁盘上两数据字间的间隔为 $1\mu\text{s}$,主存又要被 CPU 占有一半周期时间,试计算这种情况下主存周期最少应是多少?

解: (1) 根据程序查询方式的流程图可见,程序查询方式至少需要 5 条指令才能完成一个数据的传送,假定每条指令执行时间为 $1\mu\text{s}$,则两个数据字之间的最短时间间隔为 $5\mu\text{s}$ (假设每条指令 $1\mu\text{s}$)。

(2) 采用中断方式传送,这个时间并不会缩短,因为程序切换时有许多辅助操作要执行。更好的传送方式是 DMA 方式。

(3) 在 DMA 方式下,假设磁盘上两数据字间的间隔为 $1\mu\text{s}$,主存又要被 CPU 占有一半周期时间,需要采用存储器分时法,此时主存周期最少应是原来的一半,即 $0.5\mu\text{s}$ 。

25. 磁盘机采用 DMA 方式与主机通信,若主存周期为 $1\mu\text{s}$,能否满足传送速率为 1MB/s 的磁盘机的要求? 此时 CPU 处于什么状态? 若要求主存有一半时间允许 CPU 访问,该如何办?

解: 刚好能满足磁盘机的要求,但此时 CPU 只能处于停止访问主存方法。若要求主存有一半时间允许 CPU 访问,则主存的存取周期必须提高到 $0.5\mu\text{s}$ 。

26. 什么是通道控制方式? 通道和设备控制器各有哪些基本功能?

解: 通道是一种高级的 I/O 控制部件,它在一定的硬件基础上利用软件手段实现对 I/O 的控制和传送,更多地免去了 CPU 的介入,从而使主机和外设的并行工作程度更高。

通道的基本功能为:

- ① 接受 CPU 的 I/O 指令,按指令要求与指定的外设进行联系;
- ② 从主存取出属于该通道程序的通道指令,经译码后向设备控制器和设备发送各种命令;
- ③ 实施主存和外设间的数据传送,如为主存或外设装配和拆卸信息,提供数据中间缓存的空间以及指示数据存放的主存地址和传送的数据量;
- ④ 从外设获得设备的状态信息,形成并保存通道本身的状态信息,根据要求将这些状态信息送到主存的指定单元,供 CPU 使用;
- ⑤ 将外设的中断请求和通道本身的中断请求按次序及时报告 CPU。

设备控制器的基本功能为:

- ① 从通道接收控制信号,控制外部设备完成所要求的操作;
- ② 向通道反馈外部设备的状态;
- ③ 将外部设备的各种不同信号转换为通道能识别的标准信号。

27. 通道有哪些基本类型? 各有何特点?

解: 通道可分为 3 种基本类型:字节多路通道、选择通道和数组多路通道。

字节多路通道是一种简单的共享通道,用于连接与管理多台低速设备,以字节交叉方式传送信息。

选择通道也可以连接多个设备,但这些设备不能同时工作,在一段时间内通道只能选择一台设备进行数据传送,此时该设备可以独占整个通道。选择通道主要用于连接高速外设,以成组方式高速传送。

数组多路通道是把字节多路通道和选择通道的特点结合起来的一种通道结构。它的基本思想是:当某设备进行数据传送时,通道只为该设备服务;当设备在执行辅助操作时,通道暂时断开与这个设备的连接,挂起该设备的通道程序,去为其他设备服务。

28. 已知一个 32 位大型计算机系统具有两个选择通道和一个多路通道。每个选择通道连接两台磁盘机和两台磁带机,多路通道连接两台打印机、两台卡片输入机和 10 台 CRT 显示终端。假设这些设备的传送速率分别为

磁盘机	800KB/s
磁盘机	200KB/s
打印机	6.6KB/s
卡片输入机	1.2KB/s
CRT 显示终端	1KB/s

求该计算机系统的最大 I/O 传送速率。

解: 由于两个选择通道所连接的设备相同,只要计算其中一个通道的通道传输率即可。因为磁盘机的传输率大于磁带机,所以此类型通道的通道传输率为:

$$\text{选择通道传输率} = \max\{800, 200\} = 800 \text{KB/s}$$

字节多路通道的最大传输率是通道上所有设备的数据传输率之和。即:

$$\text{字节多路通道传输率} = 6.6 \times 2 + 1.2 \times 2 + 1 \times 10 = 25.6 \text{KB/s}$$

$$\begin{aligned} \text{计算机系统最大 I/O 数据传输率} &= 2 \times \text{选择通道传输率} + \text{字节多路通道传输率} \\ &= 800 \times 2 + 25.6 = 1625.6 \text{KB/s} \end{aligned}$$

29. 某计算机 I/O 系统中,接有一个字节多路通道和一个选择通道。字节多路通道包括 3 个子通道。其中:0 号子通道上接有两台打印机(传送率为 5KB/s);1 号子通道上接有 3 台卡片输入机(传送率为 1.5KB/s);2 号子通道上接 8 台显示器(传送率为 1KB/s)。选择通道上接两台磁盘机(传送率为 800KB/s);5 台磁带机(传送率为 250KB/s),求 I/O 系统的实际最大流量。若 I/O 系统的极限容量为 822KB/s,问能否满足所连接设备流量的要求?

$$\text{解: 字节多路通道传输率} = 5 \times 2 + 1.5 \times 3 + 1 \times 8 = 22.5 \text{KB/s}$$

$$\text{选择通道传输率} = \max\{800, 500\} = 800 \text{KB/s}$$

$$\text{计算机系统最大 I/O 数据传输率} = \text{选择通道传输率} + \text{字节多路通道传输率} = 800 + 22.5 = 822.5 \text{KB/s}$$

不能满足所连接设备流量的要求。

30. 试概括通道控制方式和 DMA 方式的异同点。

解: DMA 和通道控制方式最基本的相同点是从 CPU 中接管外设与主存交换数据过程的控制权,使外设能与主机并行工作。它们之间主要的不同之处在于:

① DMA 与通道的工作原理不同。DMA 通过专门设计的硬件控制逻辑来控制数据交换的过程;而通道则是一个具有特殊功能的处理器,它具有自己的指令和程序,通过执行通道程序来控制数据交换的过程。

② DMA 与通道的功能不同。通道是在 DMA 的基础上发展起来的,因此通道要比 DMA 的功能更强。

③ DMA 与通道所控制的外设类型不同。DMA 只能控制速度较快、类型单一的外设;而通道则可以支持多种类型的外设。

31. 什么是通道指令? 通道指令的结构如何? 它与 CPU 指令有何区别? 它们的执行过程相同吗?

解: 通道指令也就是通道命令字 CCW,用它来编制通道程序,并由管理程序存放在主存的任何地方。通道指令的格式因计算机不同而异,通常有命令码、数据地址、传送字节计数和标志码几部分。通道指令与 CPU 指令不同,通道指令不由 CPU 执行,它不出现在指令系统中。通道指令和 CPU 指令都存放在主存中,但通道指令由通道来执行,CPU 指令由 CPU 来执行,两者的执行过程是不相同的。

32. 简述通道操作的基本过程。

解: 通道完成一次数据传输的主要过程分为以下 3 步:

① 在用户程序中使用访管指令进入管理程序,由 CPU 通过管理程序组织一个通道程序,并启动通道。

② 通道执行 CPU 为它组织的通道程序,完成指定的数据输入输出工作。

③ 通道程序结束后向 CPU 发中断请求。CPU 响应这个中断请求后,第二次进入操作系统,调用管理程序对中断请求进行处理。

33. 在通道控制方式下,I/O 操作由通道控制,以达到 CPU 和 I/O 设备的并行操作,试问:

(1) 当通道正在进行 I/O 操作时,CPU 能否响应其他中断请求?

(2) 若 CPU 能响应其他中断请求,是否会影响正在进行的 I/O 操作?

解: (1) 当通道正在进行 I/O 操作时,CPU 可以响应其他的中断请求。

(2) 若 CPU 能响应其他中断请求,不会影响正在进行的 I/O 操作,因为 I/O 操作是由通道来控制的,与 CPU 没有关系。

34. 为什么要设立总线仲裁机构? 集中式总线控制常用哪些方式? 它们各有什么优缺点?

解: 由于总线是公共的,为了保证同一时刻只有一个申请者使用总线,总线控制机构中设置有总线判优和仲裁控制逻辑,即按照一定的优先次序来决定哪个部件首先使用总

线,只有获得总线使用权的部件,才能开始数据传送。

集中式总线控制有3种常见的优先权仲裁方式:链式查询方式、计数器定时查询方式和独立请求方式。它们各自的优缺点是:

链式查询的优点是只用很少几根线就能按一定的优先次序来实现总线控制,并易于扩充。缺点是对查询链的故障很敏感,查询的优先级是固定的。

计数器定时查询方式可以方便地改变优先次序,增加系统的灵活性,但控制线数稍多。

独立请求方式的优点是响应时间快,然而这是以增加控制线数和硬件电路为代价的。此方式对优先次序的控制也是相当灵活的,它可以预先固定,也可以通过程序来改变优先次序。

35. 总线的同步通信和异步通信有何不同? 试举例说明一次全互锁异步应答的通信情况。

解: 总线的同步通信是指系统采用一个统一的时钟信号来协调发送和接收双方的传送定时关系。时钟产生相等的时间间隔,每个间隔构成一个总线周期。总线的异步通信没有公用的时钟,也没有固定的时间间隔,完全依靠传送双方相互制约的“握手”信号来实现定时控制。

全互锁异步应答的通信过程为:“请求”信号的来到导致“回答”信号的来到,“请求”信号的撤销取决于“回答”信号的来到,而“请求”信号的撤销又导致“回答”信号的撤销。

参 考 文 献

- 1 蒋本珊.计算机组成原理.北京:清华大学出版社,2004
- 2 蒋本珊.计算机组成原理学习指导与习题解析.北京:清华大学出版社,2005
- 3 蒋本珊等.计算机专业硕士研究生入学考试重点课程辅导——计算机组成原理·离散数学·操作系统.北京:人民邮电出版社,2002